C2 (Codage, Cryptographie) et Biométrie

Travail effectué avec Julien Bringer, Bruno Kindarji (Sagem Sécurité)

avec Malika Izabachène, David Pointcheval, Qiang Tang, Sébastien Zimmer (ENS) lors du projet ANR 2005 BACH

avec Gérard Cohen (TELECOM ParisTech)

et Gilles Zémor (Inst. Math. Bordeaux)

Hervé Chabanne

Sagem Sécurité

Pôle Recherche en sécurité et cryptographie

Journées Codage et Cryptographie 2008 (Carcans)

Ordre de la présentation

- Motivation : introduction de la problématique
  - Reconnaissance biométrique
  - Respect de la vie privée
- Solutions
  - Secure sketches
  - PIR étendu
- Conclusion

Ordre de la présentation

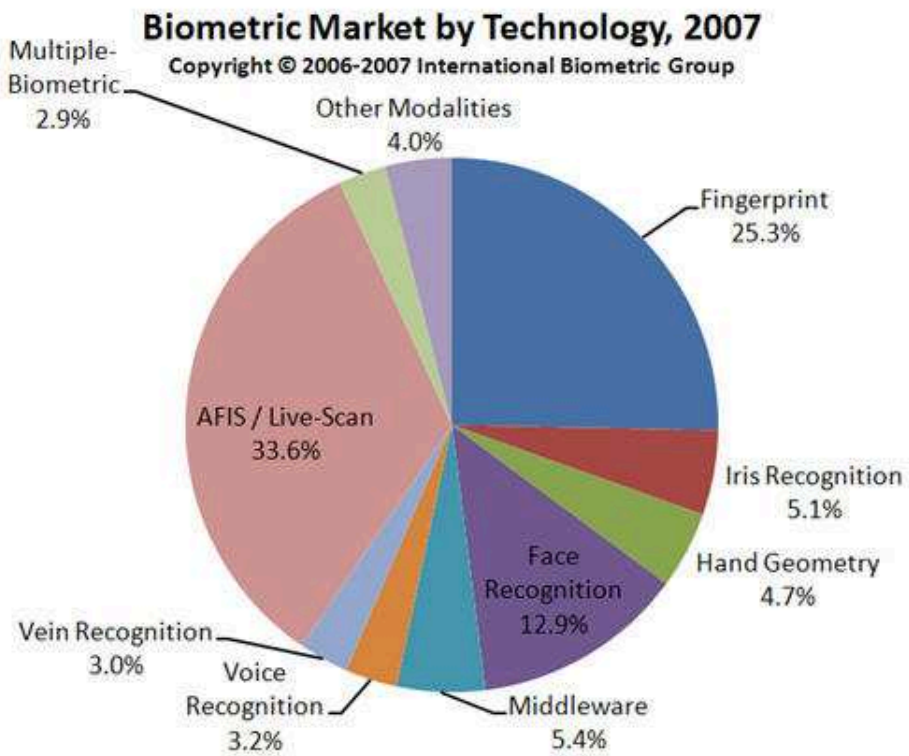
- Reconnaissance biométrique
- Respect de la vie privée
- Secure sketches
- PIR étendu
- Conclusion

Introduction

- Partie entière d'un individu, persistante dans le temps : permet une identification parmi une large population

Introduction

- Partie entière d'un individu, persistante dans le temps : permet une identification parmi une large population



Introduction

- Partie entière d'un individu, persistante dans le temps : permet une identification parmi une large population
- Le 3ième facteur (ce que je suis) : facilement accessible, n'est pas choisie, n'est pas transférable

Principe de la reconnaissance biométrique

Soit W un trait biométrique, les différentes acquisitions de ce trait sont notées $w_0, w_1, w_2 \ldots \leftarrow W$

1. Enrôlement : Une acquisition de référence $w \leftarrow W$ est conservée (dans une base de données $\mathcal{DB}$)
2. Vérification : Une nouvelle acquisition $w' \leftarrow W$ est comparée à w

Principe de la reconnaissance biométrique

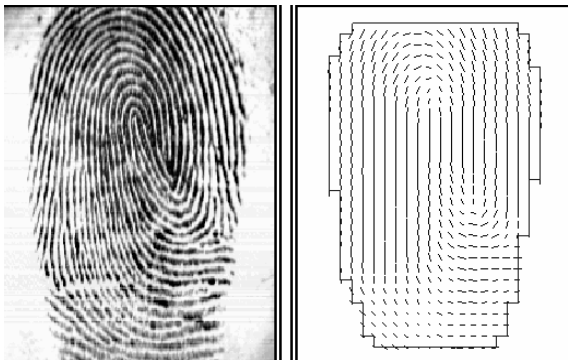
Soit W un trait biométrique, les différentes acquisitions de ce trait sont notées $w_0, w_1, w_2 \ldots \leftarrow W$

1. Enrôlement : Une acquisition de référence $w \leftarrow W$ est conservée (dans une base de données $\mathcal{DB}$)
2. Vérification : Une nouvelle acquisition $w' \leftarrow W$ est comparée à w

Remarque

Aucune confidentialité n'est requise pour que la reconnaissance fonctionne. En revanche, le capteur doit s'assurer que l'image acquise provient d'une personne vivante.
Dans la suite, nous supposerons le capteur de confiance.

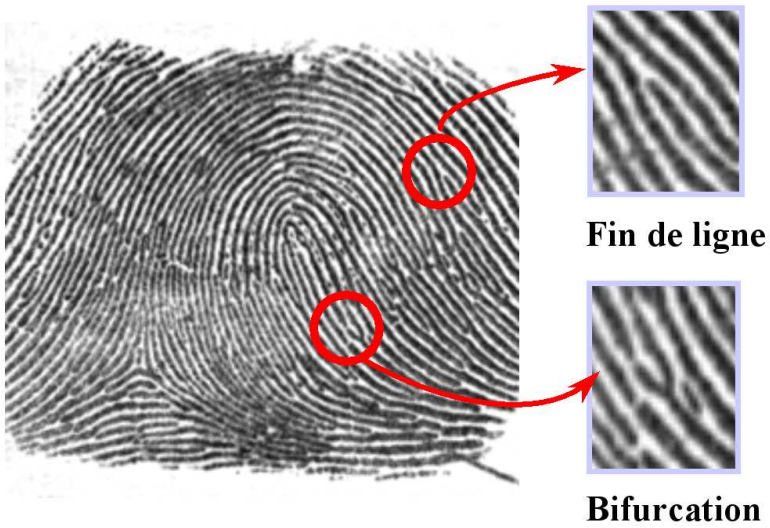
Empreintes digitales

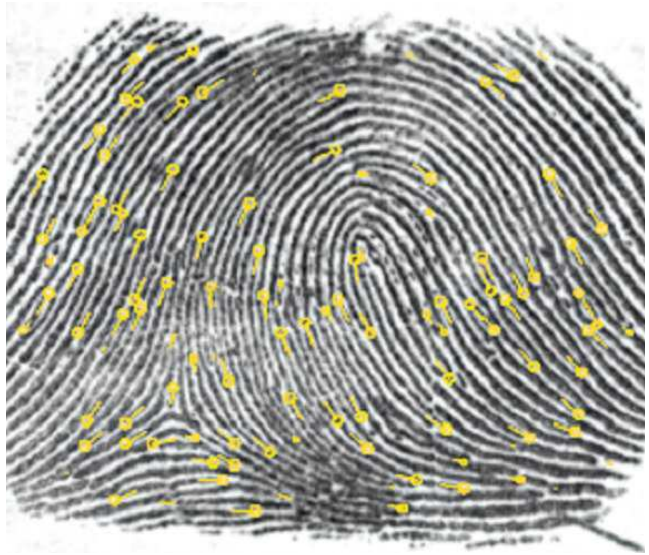


Des informations de plusieurs niveaux :

- niveau 1 (niveau global) :
  - core : point de courbure maximale
  - delta : bifurcation entre les lignes de crête
  - matrice de direction : représente en chaque point de l'image, la direction générale des lignes de l'empreinte

Empreintes digitales

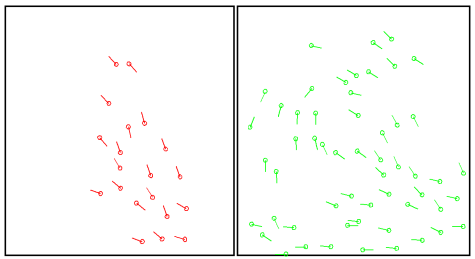


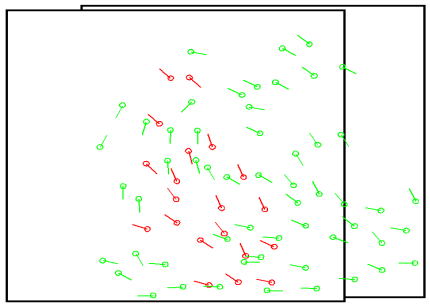


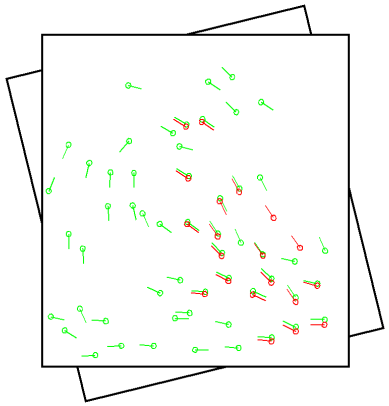
Des informations de plusieurs niveaux :

- niveau 2 (niveau local) : minuties

Empreintes digitales

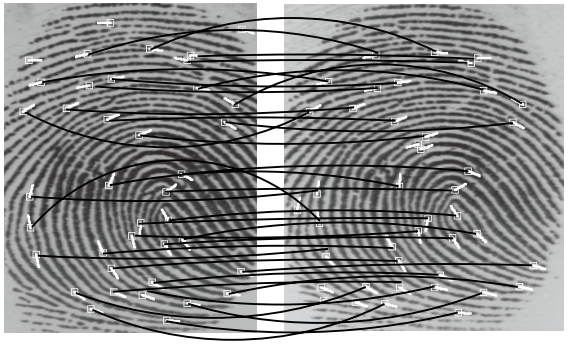


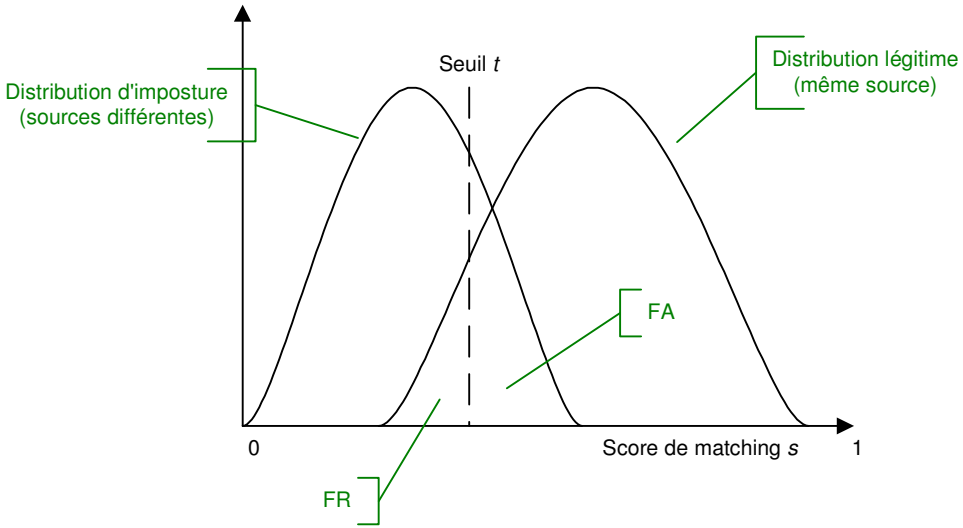




- dans un premier temps, matching global
  (positionnement du doigt sur le capteur) : rotations, translations
- dans un second temps, matching local
  (distortions type pression du doigt sur le capteur)
- obtention d'un score (une dizaine de minuties doivent se correspondre)

Matching





- Empreintes :
  - FAR $< 0,1$ %
  - FRR < 1 %
- Iris :
  - FAR : $< 10^{-4}$
  - FRR : < 1 %

Ordre de la présentation

- Reconnaissance biométrique
- Respect de la vie privée, définition du besoin cryptographique
- Secure sketches
- PIR étendu
- Conclusion

Respect de la vie privée

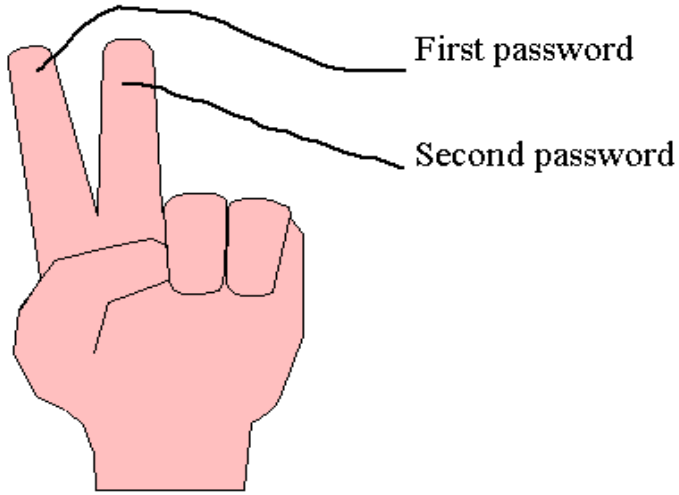
- L'appartenance d'un individu à une application doit être protégée
- Remplacement de son identité réelle par un pseudonyme
- Les acquisitions de ses traits biométriques ne doivent pas être révélées dans l'application

Respect de la vie privée

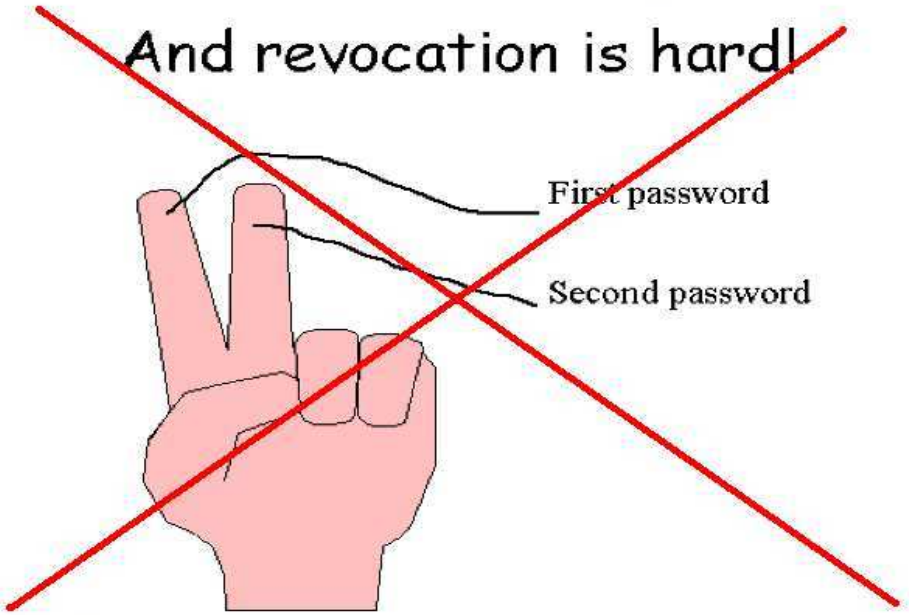
- L'appartenance d'un individu à une application doit être protégée
- Remplacement de son identité réelle par un pseudonyme
- Les acquisitions de ses traits biométriques ne doivent pas être révélées dans l'application
  - Partie entière d'un individu, persistante dans le temps : permet une identification parmi une large population

Capacité de l'adversaire

And revocation is hard!



Capacité de l'adversaire



Capacité de l'adversaire

- Un adversaire a accès à différentes acquisitions d'un trait biométrique $w_0, w_1, w_2 \ldots \leftarrow W$
- Nous voulons l'empêcher de déterminer s'il a affaire ou non à un individu donné

Problème de confidentialité

Les solutions classiques de chiffrement se heurtent à la grande variabilité des acquisitions biométriques

Chiffrement à clé publique

Trois algorithmes :

1. Gen qui, à partir d'un paramètre de sécurité 1^k, $k \in \mathbb{N}^*$, engendre la clé privée sk et la clé publique pk
2. Enc qui à un clair et à la clé publique pk associe un chiffré
3. Dec qui à un chiffré, en mettant en œuvre la clé privée sk, retrouve le clair

Chiffrement à clé publique

Trois algorithmes :

1. Gen qui, à partir d'un paramètre de sécurité 1^k, $k \in \mathbb{N}^*$, engendre la clé privée sk et la clé publique pk
2. Enc qui à un clair et à la clé publique pk associe un chiffré
3. Dec qui à un chiffré, en mettant en œuvre la clé privée sk, retrouve le clair

Propriété

Sécurité sémantique [Goldwasser-Micali 1982] : le chiffré ne révèle aucune information sur le message clair à un adversaire polynomial

Chiffrement homomorphe

Un chiffrement est homomorphe si

- à partir de Enc(a) et Enc(b), il est possible de calculer Enc($f(a, b)$)
- où f peut être, par exemple : $+, \times, \oplus$
- et sans que la clé privée soit utilisée

Chiffrement homomorphe

Un chiffrement est homomorphe si

- à partir de Enc(a) et Enc(b), il est possible de calculer Enc($f(a, b)$)
- où f peut être, par exemple : $+, \times, \oplus$
- et sans que la clé privée soit utilisée

Exemple

Chiffrement de Paillier (1999) :

- Gen *engendre*
  - $n = pq$
  - g *un entier dont l'ordre est un multiple de* n *modulo* n^2
  - $pk = (n, g)$
  - $sk = \lambda(n)$
- $\text{Enc}(m) = g^m r^n \mod n^2$

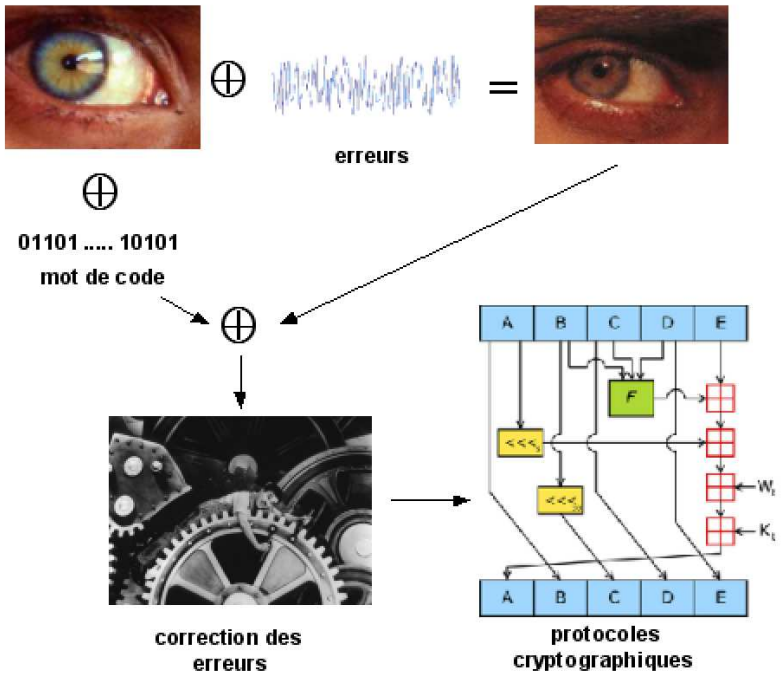
$$\text{Dec}(\text{Enc}(m) \times \text{Enc}(m') \mod n^2) = m + m' \mod n$$

$$\text{Dec}(\text{Enc}(m)^k \mod n^2) = km \mod n$$

Ordre de la présentation

- Reconnaissance biométrique
- Respect de la vie privée
- Secure sketches
  - Définition, utilisation
  - Performances
  - Sécurité, intégration des protocoles cryptographiques
- PIR étendu
- Conclusion

Principe d'utilisation



- Les données biométriques sont quantifiées
- Remplacement de l'algorithme de matching traditionnel par une correction d'erreurs grâce aux secure sketches

Secure sketches

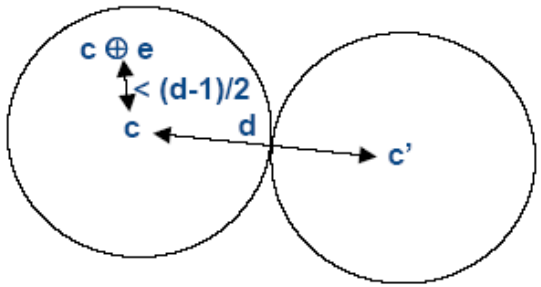
Définition (Secure sketch [Dodis-Reyzin-Smith 2004])

Soit $\mathcal{H}$ *un espace métrique avec* d *sa distance. Un* $(\mathcal{H}, m, m', t)$-secure sketch *est constitué de deux fonctions* (SS, Rec) *vérifiant :*

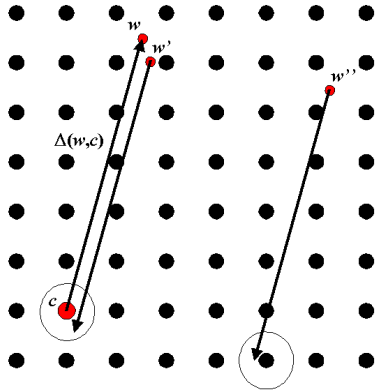
1. $\mathsf{SS} : \mathcal{H} \to \{0,1\}^*$ *est une fonction probabiliste qui à* $w \in \mathcal{H}$ *associe le Sketch de* w *:* $\mathsf{SS}(w) = \mathsf{SS}(w; X)$ *avec* X *un aléa.*
2. Rec *est une fonction déterministe prenant en entrée un élément* $w' \in \mathcal{H}$ *et un Sketch* P*, retournant* $w'' \in \mathcal{H}$ *avec* $w'' = w$ *lorsque* $P = \mathsf{SS}(w)$ *et ce, pour tous les* w' *vérifiant* $\mathrm{d}(w, w') \leq t$ *.*
3. *Pour toute variable aléatoire* W *de min-entropie* $\mathbf{H}_\infty(W) \geq m$*, la min-entropie conditionnelle de* W *sachant* $\mathsf{SS}(W)$ *vaut au moins* m'*, c'est-à-dire :* $\overline{\mathbf{H}}_\infty(W \mid \mathsf{SS}(W)) \geq m'$*.*

Codes correcteurs d'erreurs

- Soit $\mathcal{H} = \mathbb{F}_q^n$ muni de la distance de Hamming d.
- Un $[n, k, d]_q$ code est un sous-espace vectoriel de dimension k avec $d = \min_{c,c' \in C} \mathrm{d}(c, c')$
- Un tel code permet de corriger α erreurs et β effacements avec $2\alpha + \beta < d$



Construction Code-Offset



Définition (Construction Code-Offset [Juels-Wattenberg 1999])

Soit $\mathcal{H} = \mathbb{F}_q^n$ muni de la distance de Hamming. Soit C un $[n, k, d]_q$ code. Le secure sketch $(\mathsf{SS}_C, \mathsf{Rec}_C)$ est défini par :

- *la fonction SS_C prend w comme entrée et retourne le Sketch $P = c - w$, où c est un mot du code C pris au hasard.*
- *la fonction Rec_C prenant w' et P comme entrées, décode $w' + P = c + (w' - w) = c + e$ comme un mot de code c' et retourne alors $c' - P$.*

Binarisation de l'empreinte (1/2) (Tuyls et al. 2005)

M images sont acquises pour le trait biométrique d'un individu
une base de données pour N individus faite de NM vecteurs réels $(X_{i,j})_{i=1..N,j=1..M}$

Traitement statistique :

- Moyenne par individu : $\forall i \in \{1,\ldots,N\}$, $\mu_i = \frac{1}{M}\sum_{j=1}^{M} X_{i,j}$.
- Moyenne générale : $\mu = \frac{1}{N}\sum_{k=1}^{N} \mu_i$.

Binarisation de l'empreinte (2/2)

Quantification :
Pour chaque coordonnée $X_{i,j}$,le vecteur binaire $Q(X_{i,j})$ est calculé :

$$\forall t \in \{1, \ldots, L\},\ (Q(X_{i,j}))_t = 0 \text{ si } (X_{i,j})_t \leq (\mu)_t,\ 1 \text{ si } (X_{i,j})_t > (\mu)_t.$$

Binarisation de l'empreinte (2/2)

Quantification :
Pour chaque coordonnée $X_{i,j}$,le vecteur binaire $Q(X_{i,j})$ est calculé :

$$\forall t \in \{1, \ldots, L\},\ (Q(X_{i,j}))_t = 0 \text{ si } (X_{i,j})_t \leq (\mu)_t,\ 1 \text{ si } (X_{i,j})_t > (\mu)_t.$$

Sélection de bits fiables :
Pour chaque individu i_0, les coordonnées de $Q(X_{i_0,j})$ donnant le plus souvent le même bit pour $Q(X_{i_0,j})$ sont choisies.
Les indices des coordonnées retenues sont conservés dans un vecteur, appelé *Helper Data* P_{1,i_0}.
Ce vecteur est gardé avec le *Secure Sketch* $P_{2,i_0} = \mathrm{SS}_C(w) = c_{i_0} \oplus w$.

Ordre de la présentation

- Reconnaissance biométrique
- Respect de la vie privée
- Secure sketches
  - Définition, utilisation
  - Performances
  - Sécurité, intégration des protocoles cryptographiques
- PIR étendu
- Conclusion

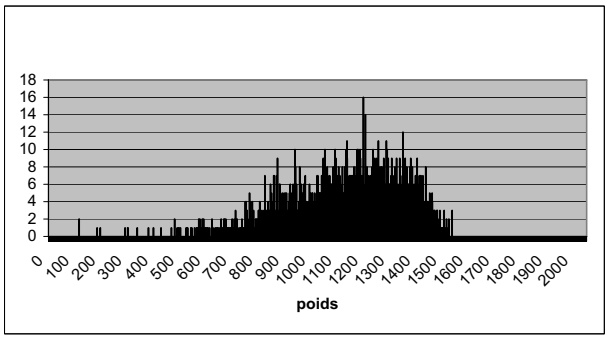
Codes adaptés aux secure sketches : un premier résultat théorique [BCCKZ BTAS 2007]

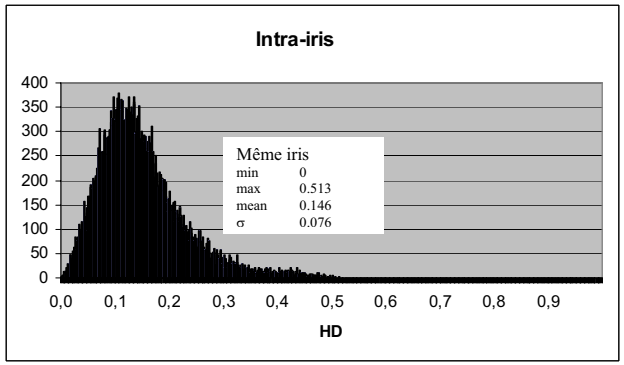
Théorème

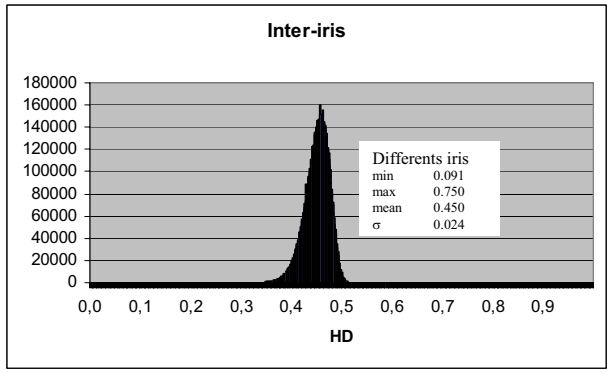
Soient $k \in \mathbb{N}^*$ et C un code binaire de longueur n possédant 2^k mots. Soit m un mot de C entaché de α erreurs et β effacements. Supposons que C soit optimal par rapport à n et k et que nous décodons au maximum de vraisemblance.
Si $\frac{\alpha}{n-\beta} > \theta$ alors la probabilité que m soit décodable est négligeable lorsque n est grand, où θ est tel que la sphère de Hamming dans $\mathbb{F}_2^{n-\beta}$ de rayon $(n-\beta)\theta$ contienne $2^{n-\beta-k}$ éléments , avec la sphère définie comme l'ensemble $\{x \in \mathbb{F}_2^{n-\beta}, \mathrm{p}(x) = (n-\beta)\theta\}$.

Conséquences

Etude de la base d'iris du NIST ICE, $n = 2048$

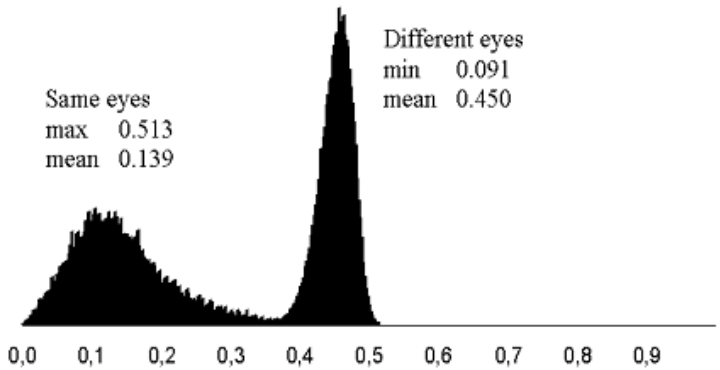






Conséquences

Etude de la base d'iris du NIST ICE, $n = 2048$



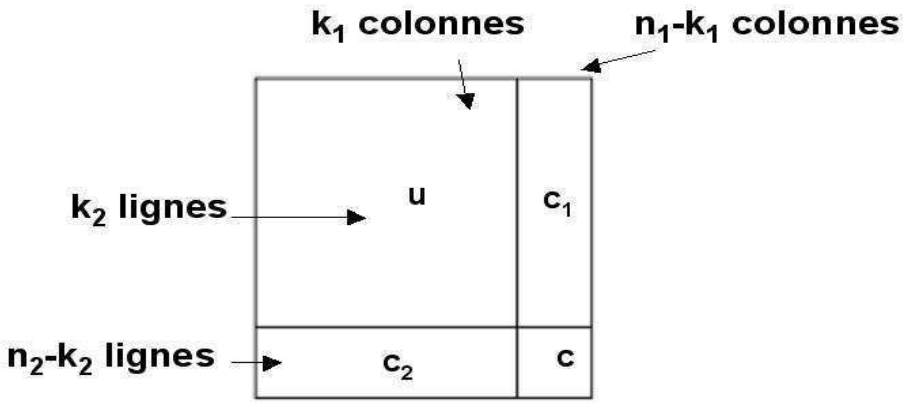
Conséquences

Etude de la base d'iris du NIST ICE, $n = 2048$

- $k = 42$, FRR $> 2,49$ %
- $k = 64$, FRR $> 3,76$ %
- $k = 80$, FRR $> 4,87$ %
- $k = 128$, FRR $> 9,10$ %

Code et décodage associé

- Faisant suite à [Hao-Anderson-Daugman 2006]
- Canal " biométrique " mal connu
- Choix d'un code produit $C = RM(1,6) \otimes RM(1,5)$ de longueur 2048 et de dimension 42
- Décodage exhaustif de chacun des RM
- Décodage de C avec l'algorithme itératif min-sum
- Entrelacement aléatoire
- FRR $= 5,62\ \%\ > 2,49\ \%$ (1 %), FAR $= 0,0006\ \%$ (10^{-4})



Ordre de la présentation

- Reconnaissance biométrique
- Respect de la vie privée
- Secure sketches
  - Définition, utilisation
  - Performances
  - Sécurité, intégration des protocoles cryptographiques
- PIR étendu
- Conclusion

Protocole d'authentification de Juels et Wattenberg

1. A l'enrôlement, conserver $P = \mathrm{SS}_{\mathrm{C}}(w) = c - w$ et $H(c)$ où c est pris au hasard dans C et où H est une fonction de hachage cryptographique.
2. Lorsqu'un individu veut s'authentifier en présentant w', décoder $w' + P = c + (w' - w)$ en c' et vérifier si $H(c') = H(c)$.

Sécurité des secure sketches

- Sécurité mal établie en pratique
- Ne protège pas contre des vérifications a posteriori
  (i.e. test d'appartenance d'un individu à une base donnée)
  [BCD IEEE-IT 2006]

Authentification sur le principe de Juels et Wattenberg avec des données chiffrées

- Utilisation de cryptosystèmes homomorphes
- Par exemple, chiffrement de Goldwasser-Micali :

$$\sqsubset b \sqsupset \times \sqsubset b' \sqsupset = \sqsubset b \oplus b' \sqsupset$$

en notant $\sqsubset b \sqsupset = y^2 x^b \mod n$ où y est tiré au hasard

Authentification sur le principe de Juels et Wattenberg avec des données chiffrées

- Utilisation de cryptosystèmes homomorphes
- Par exemple, chiffrement de Goldwasser-Micali :

$$\sqsubset b \sqsupset \times \sqsubset b' \sqsupset = \sqsubset b \oplus b' \sqsupset$$

en notant $\sqsubset b \sqsupset = y^2 x^b \mod n$ où y est tiré au hasard

1. A l'enrôlement, calculer $P = \mathrm{SS}_C(w) = c \oplus w$ et $H(c)$ où c est pris au hasard dans C et où H est une fonction de hachage cryptographique. Alice conserve $\sqsubset P \sqsupset$ et le serveur stocke $H(c)$ dans une base de données

Authentification sur le principe de Juels et Wattenberg avec des données chiffrées

- Utilisation de cryptosystèmes homomorphes
- Par exemple, chiffrement de Goldwasser-Micali :

$$\sqsubset b \sqsupset \times \sqsubset b' \sqsupset = \sqsubset b \oplus b' \sqsupset$$

en notant $\sqsubset b \sqsupset = y^2 x^b \mod n$ où y est tiré au hasard

1. A l'enrôlement, calculer $P = \mathrm{SS}_C(w) = c \oplus w$ et $H(c)$ où c est pris au hasard dans C et où H est une fonction de hachage cryptographique. Alice conserve $\sqsubset P \sqsupset$ et le serveur stocke $H(c)$ dans une base de données
2. Lorsqu'Alice veut s'authentifier grâce à w', elle calcule $\sqsubset P \sqsupset \times \sqsubset w' \sqsupset = \sqsubset c \oplus w \oplus w' \sqsupset = \sqsubset c \oplus e \sqsupset = Z$.
   Le serveur déchiffre Z, décode $w' \oplus P$ en c' et vérifie si $H(c') = H(c)$

Intégration du matching dans des protocoles d'authentification biométrique

- Séparation nette entre données biométriques et données temporaires
- Stockage des données biométriques dans une base de données [BCIPTZ ACISP 2007]

Private Information Retrieval

Définition (PIR [Chor-Kushilevitz-Goldreich-Sudan 1998])

Un protocole de PIR permet à un utilisateur d'interroger une base de données pour obtenir un bit. En notant i l'indice du bit demandé, le protocole doit respecter les propriétés suivantes :

- *le protocole retourne le bit i*
- *respect de la vie privée : la base de données ne doit rien apprendre sur le bit i*

Définition (PBR [Chor-Kushilevitz-Goldreich-Sudan 1998])

Un protocole PBR permet d'obtenir les données par bloc.

Intégration dans un protocole de PIR (1/2) [BC Africacrypt 2008]

Système composé

- de M utilisateurs,
- d'un serveur d'authentification fait :
  - d'une base de données $\mathcal{DB}$ contenant les l bits des secure sketches chiffrés par le cryptosystème de Goldwasser-Micali $a_{k,j} = \sqsubset P_{k,j} \sqsupset = \sqsubset (c_k \oplus w_k)_j \sqsupset$, $k = 1, \ldots, M$,
  - d'un fournisseur d'accès gérant les demandes d'authentification qui a accès aux hachés cryptographiques correspondants $H(c_k)$,
  - d'un Hardware Security Module gardant les clés du système.

Intégration dans un protocole de PIR (2/2)

En notant

- $\sqsubset s \sqsupset$ le chiffrement de s par le cryptosystème de Goldwasser-Micali
- $\prec t \succ$ le chiffrement de t par le cryptosystème de Paillier

Lorsqu'un individu i veut s'authentifier :

1. La base $\mathcal{DB}$ calcule suivant un schéma de PIR linéaire classique :

$$\prec \sqsubset P_{i,j} \sqsupset \succ = \prod_{k=1,\ldots,M} \prec \delta_k \succ^{a_{k,j}}$$

avec $\delta_k = 1$ si $k = i$ et 0 sinon

Intégration dans un protocole de PIR (2/2)

En notant

- $\sqsubset s \sqsupset$ le chiffrement de s par le cryptosystème de Goldwasser-Micali
- $\prec t \succ$ le chiffrement de t par le cryptosystème de Paillier

Lorsqu'un individu i veut s'authentifier :

1. La base $\mathcal{DB}$ calcule suivant un schéma de PIR linéaire classique :

$$\prec \sqsubset P_{i,j} \sqsupset \succ = \prod_{k=1,\ldots,M} \prec \delta_k \succ^{a_{k,j}}$$

   avec $\delta_k = 1$ si $k = i$ et 0 sinon
2. Le fournisseur d'accès modifie les chiffrés par

$$\prec \sqsubset P_{i,j} \sqsupset \succ^{\sqsubset w'_j \sqsupset} = \prec \sqsubset P_{i,j} \oplus w'_j \sqsupset \succ,\ j = 0, \ldots, l-1$$

Propriété

$$\prec \sqsubset s \sqsupset \succ^{\sqsubset w \sqsupset} = \prec \sqsubset s \sqsupset \times \sqsubset w \sqsupset \succ = \prec \sqsubset s \oplus w \sqsupset \succ$$

Intégration dans un protocole de PIR (2/2)

En notant

- $\sqsubset s \sqsupset$ le chiffrement de s par le cryptosystème de Goldwasser-Micali
- $\prec t \succ$ le chiffrement de t par le cryptosystème de Paillier

Lorsqu'un individu i veut s'authentifier :

1. La base $\mathcal{DB}$ calcule suivant un schéma de PIR linéaire classique :

$$\prec\sqsubset P_{i,j} \sqsupset\succ = \prod_{k=1,\ldots,M} \prec \delta_k \succ^{a_{k,j}}$$

   avec $\delta_k = 1$ si $k = i$ et 0 sinon
2. Le fournisseur d'accès modifie les chiffrés par

$$\prec\sqsubset P_{i,j} \sqsupset\succ^{\sqsubset w'_j \sqsupset} = \prec\sqsubset P_{i,j} \oplus w'_j \sqsupset\succ,\ j = 0, \ldots, l-1$$

Remarque

Applicable au protocole de PIR de Lipmaa en remplaçant le cryptosystème de Paillier par sa généralisation par Damgard et Jurik

Ordre de la présentation

- Reconnaissance biométrique
- Respect de la vie privée
- Secure sketches
- PIR étendu
- Conclusion

EPIR : PIR Etendu [BCPT CANS 2007]

- Dans nos protocoles d'authentification, nous n'avons pas besoin de la valeur de référence mais seulement d'un résultat sur la valeur

EPIR : PIR Etendu [BCPT CANS 2007]

- Base de données $\mathcal{DB}$ contient les données biométriques chiffrées : $(R_1, \ldots, R_n)$
- A un utilisateur $\mathcal{U}$ sont associés un indice i et un champ R_i
- $\mathcal{U}$ veut effectuer une évaluation d'une fonction simple $f(X, R_i)$,

EPIR : PIR Etendu [BCPT CANS 2007]

- Base de données $\mathcal{DB}$ contient les données biométriques chiffrées : $(R_1, \ldots, R_n)$
- A un utilisateur $\mathcal{U}$ sont associés un indice i et un champ R_i
- $\mathcal{U}$ veut effectuer une évaluation d'une fonction simple $f(X, R_i)$, par exemple : $f(X, R_i) =$
  - $(X \stackrel{?}{=} R_i)$
  - $\mathrm{d}(X, R_i)$

EPIR : PIR Etendu [BCPT CANS 2007]

- Base de données $\mathcal{DB}$ contient les données biométriques chiffrées : $(R_1, \ldots, R_n)$
- A un utilisateur $\mathcal{U}$ sont associés un indice i et un champ R_i
- $\mathcal{U}$ veut effectuer une évaluation d'une fonction simple $f(X, R_i)$, par exemple : $f(X, R_i) =$
  - $(X \stackrel{?}{=} R_i)$
  - $\mathrm{d}(X, R_i)$
- Propriétés attendues :
  - Permet l'évaluation de f souhaitée
  - Respect de la vie privée de $\mathcal{U}$
  - Respect de la base $\mathcal{DB}$

Respect de la vie privée de $\mathcal{U}$

La fonction f est fixée.
L'adversaire $\mathcal{A}$ joue le rôle de la base de données et cherche à apprendre des informations sur les requêtes de $\mathcal{U}$.
Il ne doit pas avoir d'avantage significatif au terme du jeu d'indistinguabilité suivant :

- $\mathcal{A}_1$ *engendre une base de données* $(R_1, \ldots, R_n)$
- $\mathcal{A}_2$ *attend que le challenger ait formulé ses requêtes.* $\mathcal{A}_2$ *retourne alors* (i_0, i_1, X_0, X_1)
- *Le challenger choisit* $b \in \{0, 1\}$ *au hasard et formule à* $\mathcal{A}_3$ *une requête concernant* (i_b, X_b)
- $\mathcal{A}_4$ *laisse formuler au challenger d'autres requêtes.* $\mathcal{A}_4$ *doit alors proposer un* b'

Respect de la base $\mathcal{DB}$

La fonction f est fixée.
L'adversaire $\mathcal{A}$ joue le rôle de l'utilisateur i et essaye de distinguer un dialogue entre la vraie base de données $\mathcal{DB}$ et un simulateur.
Le simulateur connaît seulement les réponses $f(X, R_i)$ aux requêtes (i, X).
L'adversaire ne doit pas avoir d'avantage significatif au terme du jeu suivant :

- *Le challenger choisit* $b \in \{0, 1\}$ *au hasard.*
  Si $b = 0$ *alors* $\mathcal{A}$ *interagira avec la vraie base données*
  Si $b = 1$ *alors* $\mathcal{A}$ *interagira avec le simulateur*
- $\mathcal{A}_1$ *engendre la base de données* $(R_1, \ldots, R_n)$
- $\mathcal{A}_2$ *formule des requêtes (il ne sait pas s'il interagit avec* $\mathcal{DB}$ *ou le simulateur).* $\mathcal{A}_2$ *doit alors proposer un* b'

EPIR, premier exemple : l'égalité

Variante additive du cryptosystème d'ElGamal :

- $sk = \alpha$
- $pk = y = g^{\alpha} \mod p$
- $\mathcal{E}(X) = \mathcal{E}(X, r) = (g^r \mod p, y^r g^X \mod p)$

$\mathcal{U}$ veut évaluer $f(R_i, X) = (R_i \stackrel{?}{=} X)$

EPIR, premier exemple : l'égalité

Variante additive du cryptosystème d'ElGamal :

- $sk = \alpha$
- $pk = y = g^{\alpha} \mod p$
- $\mathcal{E}(X) = \mathcal{E}(X, r) = (g^r \mod p, y^r g^X \mod p)$

$\mathcal{U}$ veut évaluer $f(R_i, X) = (R_i \stackrel{?}{=} X)$

1. $\mathcal{U}$ envoie $c = \mathcal{E}(i||X)$
2. $\mathcal{DB}$ calcule une image temporaire de ses données

$$C_j = (c/\mathcal{E}(j||R_j))^{r_j} \mod p = \mathcal{E}((i||X - j||R_j) \times r_j)$$

3. $\mathcal{U}$ et $\mathcal{DB}$ engagent un protocole de PIR pour obtenir C_i

   $\mathcal{U}$ déchiffre C_i et en déduit $X \stackrel{?}{=} R_i$

Analyse de sécurité

- si le protocole de PIR sous-jacent se déroule normalement
  alors le protocole EPIR se déroulera également normalement
- si le protocole de PIR sous-jacent respecte la vie privée de $\mathcal{U}$
  et si le problème DDH est difficile
  alors le protocole EPIR respectera aussi la vie privée de $\mathcal{U}$
- tous les champs de l'image temporaire de la base $\mathcal{DB}$ sont aléatoires
  sauf le i-ème

Chiffrement doublement homomorphe de Boneh, Goh et Nissim (BGN, 2005)

Définition (Chiffrement BGN)

- Gen *engendre*
  - $n = q_1 q_2$,
  - $\mathbb{G}$, $\mathbb{G}_1$ *deux groupes cycliques d'ordre* n
  - $\hat{e} : \mathbb{G} \times \mathbb{G} \to \mathbb{G}_1$ *une forme bilinéaire non dégénérée*
  - g *et* u *deux générateurs de* $\mathbb{G}$, $h = u^{q_2}$
  - *la clé privée* $sk = q_1$
  - *la clé publique* $pk = (n, \mathbb{G}, \mathbb{G}_1, \hat{e}, g, h)$
- Enc *chiffre* $m \in \{0, \ldots, T\} \subseteq \mathbb{Z}_{q_2}$ *avec* pk *comme* $\mathrm{Enc}(m) = g^m h^r = c \in \mathbb{G}$, $r \in \mathbb{Z}_n^*$
- Enc *déchiffre* c *avec* sk *en calculant* $c^{q_1} = (g^{q_1})^m$ *puis en prenant le logarithme discret de* c^{q_1} *en base* g^{q_1}

EPIR, second exemple : la distance de Hamming

- Propriétés du cryptosystème de BGN
  - additivement homomorphe
  - multiplicativement homomorphe
    (une fois seulement, utilise $\hat{e} : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_1$)
  - permet l'évaluation de polynômes multivariés de degré total 2
- Nous permet un EPIR avec $f(R_i, X) = \sum_{k=0}^{l-1} w_k \times (R_i^{(k)} \oplus X^{(k)})$
- En reprenant la même façon de procéder que pour l'égalité
  1. $\mathcal{U}$ envoie le chiffrement de i et X
  2. $\mathcal{DB}$ calcule une image temporaire de ses données correspondante au poids de Hamming pondéré recherché
  3. $\mathcal{U}$ et $\mathcal{DB}$ engagent un protocole de PIR pour obtenir le champ idoine

Protocole d'authentification biométrique correspondant

Hypothèse

Nous supposons que pour un seuil t, la probabilité d'avoir $\mathrm{d}(w, w') \leq t$ (resp. $\mathrm{d}(w, w') > t$)
pour $w, w' \leftarrow W$ (resp. pour $w \leftarrow W$ et $w' \leftarrow W'$, $W \neq W'$)
est proche de 1.

Protocole d'authentification biométrique correspondant

Hypothèse

Nous supposons que pour un seuil t, la probabilité d'avoir $\mathrm{d}(w, w') \leq t$ (resp. $\mathrm{d}(w, w') > t$)
pour $w, w' \leftarrow W$ (resp. pour $w \leftarrow W$ et $w' \leftarrow W'$, $W \neq W'$)
est proche de 1.

Le protocole d'EPIR précédent pour calculer la distance de Hamming donne directement un protocole d'authentification biométrique

Ordre de la présentation

- Reconnaissance biométrique
- Respect de la vie privée
- Secure sketches
- PIR étendu
- Conclusion

Conclusion

- Encore du travail :
  - Quantification des données biométriques
  - Codage / Décodage pour les Secure Sketches
  - Plus grande efficacité des protocoles d'interrogation de bases de données chiffrées (aujourd'hui, le chiffrement se fait bit à bit)
- Par exemple : exposé de Bruno Kindarji ...