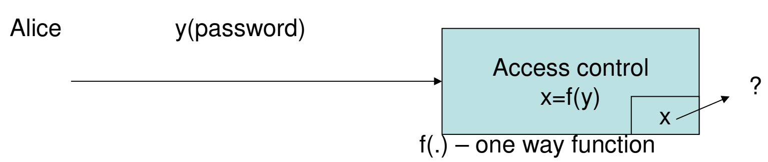
Identification and Key Distribution Based on Biometric Information

Prof. Dr. Valery Korzhik
University of Telecommunications
St. Petersburg (Russia)
-2008-

1. Identification of users



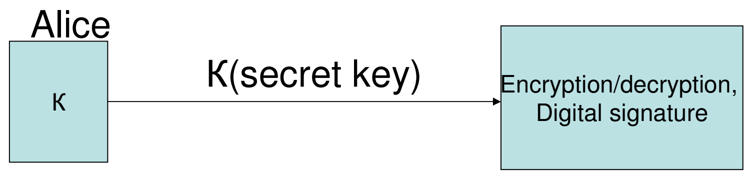
Important remark:

With the use of one-way function it is assumed that "*y*" is distributed trully randomly. Otherwise – nothing is taken for granted.

Defects of this approach:

- Good password can be forgotten by Alice,
- Storing of password in memory increases the risk of its theft,
- Short password can be easy memorized but it can be easy found by adversary

Conventional key application



Shortcoming of this approach:

The key that is storing in some memory can be stolen or erased

How can we remove this defect?

Use Alice's biometric or her psychology.

«My body is my password»[1]

+

Psychology: say Alice's preferences are: young, rich and healthy men.

Biometrics as a source of passwords and keys

The main types of biometrics:

- Palmprint verification,
- Iris biometric,
- Face recognition,
- Fingerprint system,
- Speaker recognition,
- Signature system,
- Keystroke biometrics,
- Using a small subset of values from a large universe (e.g. favorite movies),
- A combining of methods.
- Vein pattern
- Ear
- Facial thermogram
- Gait

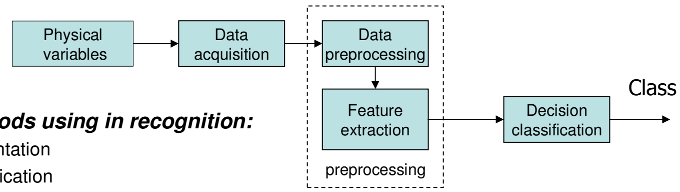
Remark:

Both hardware and software to transform human biometrics into digital form were designed by many companies [1].

Defects of biometrical approach:

Digital data producing by biometrics *are not truly random* and it is very *difficult* to *reproduce* them repeatedly.

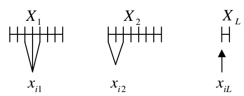
Configuration of pattern recognition system [1]



The main methods using in recognition:

- Image segmentation
- Pattern classification
- Discriminant functions
- Bayes and non-Bayes classification
- Neural networks
- Support vector machine

Model of recognition:



$X_1 \quad X_2 \quad X_L$ -set of features

$x_{i1} \quad x_{i2} \quad x_{iL}$ -the feature of i-th user

Then it is possible to store data as $h(x_{ij})$,where h – OWF (one way function)
However, if the recognition follows to the rule $i = Arg \min_i \sum_{j=1}^{L} \rho^2 (x'_{ij} - x_{ij})$,
then $h(x_{ij})$ - cannot be used.

Methods to remove defects:

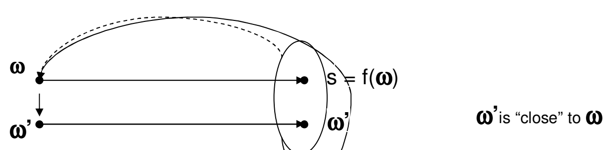
1. Recognition on biometrics. (Then the parameters have to store in secret)
2. The use of secure sketch [2] – SS
3. The use of fuzzy extractors [2] – FE

Definition and properties of SS and FE.

Non-formal definition:

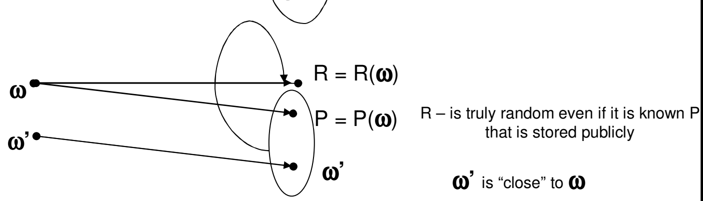
SS:

(for identification on BI)



FE:

(for key generation on BI)



Specification of the notion "ω' is close to ω":

1. Hamming metrics
$\rho_H(\omega,\omega')$ = is the number of position in which binary vectors ω and ω' are different
Example. $\omega = 10011$ $\rho_H(\omega,\omega') = 3$
$\omega' = 01010$
(This metrics is very natural for BI)
2. Set difference
$\rho_S(\omega,\omega') = \rho_S(A,A') = \frac{1}{2}|A\Delta A'|,$ Δ

where "Δ" is symmetric difference of the sets A and A'.
|B| is a cardinality of B.
Example. $U = \{1,2,3,4,5,6\}$, $A \subseteq U$, $A = 1,2,3,4$

$B \subseteq U$, $B = 3,4,5,6$, $A\Delta B = 1,2,5,6$
$\rho_S(A,B) = 2$
Psychometry: A selection of small subset from a large universe (e.g. favorite movies)

3. Edit distance
$\rho_e(\omega,\omega') = \frac{1}{2}$ (is the minimum number of omissions and insertions that are needed in order to transform ω into ω')
Example. $\omega = 1\emptyset 101\hat{1}$
$\omega' = 110111$
$\rho_e(\omega,\omega') = 1$
(This distance is very natural in recognition of handwritten text.)

3. Exact definition SS and FE:

Let us M be metrix space,

$|\mathrm{M}| = N, \quad \rho(,)$ - is given metrix

(M, m, m', t) - SS is randomized mapping

$\underset{(\omega)}{\mathrm{M}} \rightarrow \{1,0\}^*$ with a following properties:

(i) $\exists \mathrm{Rec}(...)$ such that $\omega = \mathrm{Rec}(\omega', SS(\omega))$ for all $\omega, \omega' \in \mathrm{M}, \ \rho(\omega, \omega') \le t$.

(ii) $\tilde{H}_\infty (W \mid SS(W)) \ge m'$

for any random variable W on M, having $H_\infty(W) = m$,

where $H_\infty(W) = -\log(\max_W \Pr(W))$

$H_\infty(W \mid SS(W)) = -\log(E_s\{2^{-H_\infty(W \mid SS(W)=s)}\})$

Remark: The condition (ii) makes impossible to recover W given $s = SS(W)$ unconditionally (that means that it cannot be recovered independently on computing power of opponent!)

$(\mathrm{M}, m, l, t, \varepsilon)$ - FE is determined by two procedures: (Gen, Rep):

(i) Gen – is randomized mapping

$W \in \mathrm{M} \Rightarrow R \in \{0,1\}^l$

$\searrow P$,

for which $SD(\langle R, P \rangle, \langle U_l, P \rangle) \le \varepsilon$, if $H_\infty(W) \ge m$.

(ii) Rep – is deterministic procedure $R = \mathrm{Rep}(\omega', \mathrm{P})$,

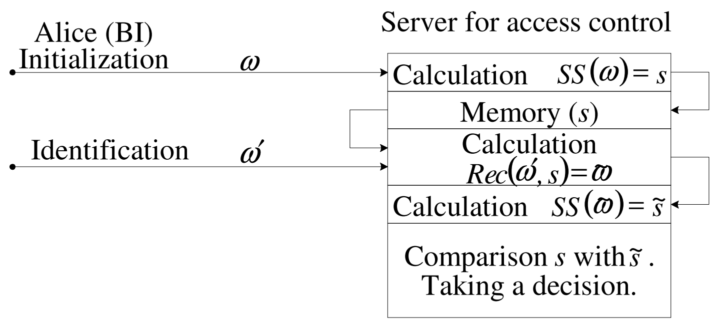
if $\rho(\omega, \omega') \le t$,

where $SD(X, Y)$ - is statistical distance between two probability distributions on X and Y, e.g.:

$$SD(X, Y) = \frac{1}{2} \sum_v |P_r(X = v) - P_r(Y = v)|.$$

Remark: The small value $SD(...)$ means that the probability distribution on $R \in \{0,1\}^l$ is close to uniform distribution (U_l) even known P, (e.g. it is close to truly random variable).

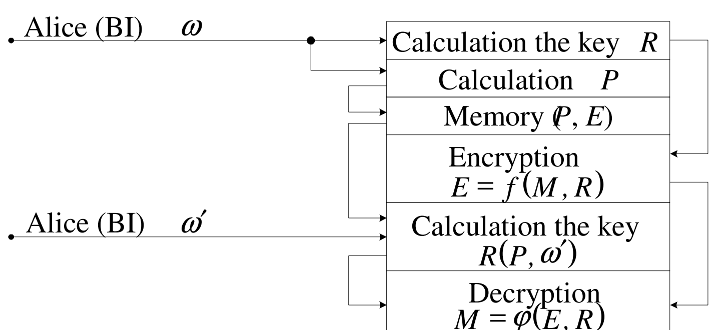
Identification based on BI using SS



Remarks:

1. Storing *s* in memory does not require a protection
2. One way functions are not needed
3. Good statistical properties (close to truly randomness) for s=SS(ω) are not provided (but they are not required)
4. It is necessary to provide a condition $H_\infty \geq m$

Key generation based on BI using FE



Remarks:

1. Key *R* is close to truly random value.
2. Storing *P*, *E* in memory does not require protection.
3. Calculation and storing *P* can be performed in a reader of BI.
4. It is necessary to protect *P* against a forgery by adversary (the use of digital signature or “robust fuzzy extractors” – see further)

4. Design of FE given SS and SE (strong extractors)

Definition SE.

SE – is randomized mapping: $\{0,1\}^n, \{0,1\}^d \rightarrow \{0,1\}^l$ such that for input strings $\omega \in \{0,1\}^n$ with arbitrary probability distribution but with min entropy at least m';

$$SD\left(SE(W, X), X; U_{l+d} \right) \le \varepsilon,$$

if $X \in \{0,1\}^d$, $Pr(X) = U_l$.

Clear demonstration of SE.

This is a generator of "good" output randomness (close to uniform distribution) presented as binary string of shorter length *l* than it's input binary string of the length *n* that has "bad" randomness given short (length *d*) truly random seed, whereas the knowledge of this seed does not affect on good output randomness.

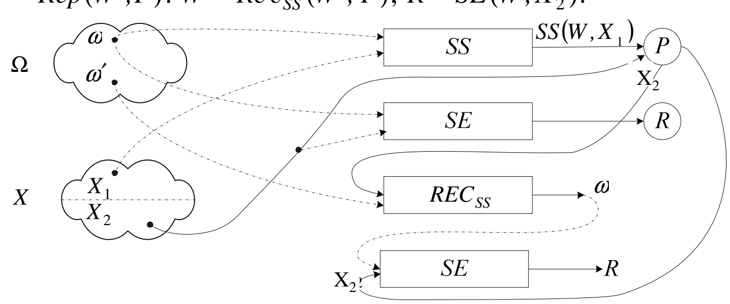
How to design FE given SE and SS?

Let us assume that there is SS(M, m, m', t) and SE(n, m', l, ε) with $l = m' - 2\log(1/\varepsilon)$.

Then the following construction (Gen, Rep) gives FE(M, m, l, t, ε):

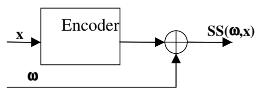
– $Gen(W; \underbrace{X_1, X_2}_{X})$: $P = SS((W; X_1), X_2)$, $R = SE(W, X_2)$.

– $Rep(W', P)$: $W = Rec_{SS}(W', P)$, $R = SE(W, X_2)$.



Conclusion: In order to design FE we have to design both SS and SE.

Construction SS for Hamming distance.
Let C is (n,k,2t+1) error correcting code (not necessary linear).
Then:
SS(ω,x)= ω⊕C(x), where x is chosen randomly and C(x) is a code word,



Rec(ω',SS(ω): ω' ⊕ SS(ω,x)= ω' ⊕ω⊕C(x)=e⊕C(x), $/e/ \le t$,
where /*e*/ is Hamming weight of *e*.
D(ω' ⊕ SS(ω,x))=D(e⊕C(x))=x, where D is a decoding procedure under the condition that C corrects at least t errors. Then
ω= SS(ω,x) ⊕C(x).
It can be proved [2] that: m'≥m-r, ∀ m,r.

Practical implementation of SS:

If C is linear code then $SS(\omega) = syn_c(\omega)$(syndrom to *w* on the code C),

e.g. $SS(\omega) = \omega H$, where H is check matrix of the code C.

In this case a randomness X is not required at all!

In fact, let us take *s=SS(ω)= ωH* and *ω'= ω⊕e,* where *e* is error pattern over the weight at most *t*. Then we have *ω'H=(ω⊕ e)H=*
= ωH⊕eH that gives relation *eH= ω'H⊕ s*. Since *C* is capable to correct all errors of the weight at most *t,* we can recover *e* on given syndrom *eH.* After that we can recover *ω* as follows: *ω = ω'⊕e.*

If $W \in U_n$, e.g.$H_\infty(W) = m = n$, then we get

FE: *R=X*,

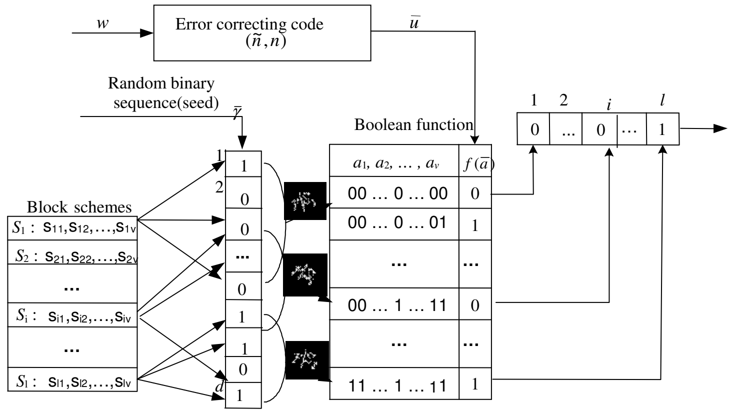
$$P = \omega \oplus C(X),$$

$$REP(\omega', P) = D(P \oplus \omega')$$

This construction does not work in general case, because if $\omega \notin U_n$, then *P* gives a leakage of information about *X=R*.

In order to design FE it is necessary to design SE.

Trevisan's extractor:



Parameters of Trevisan's extractor:

$$l = \frac{H_\infty(W)}{c}; \quad \mathrm{d} = \mathrm{O}\left(\log_2^2\left(\frac{n}{\varepsilon}\right)\frac{1}{\log_2 c}\right); \quad \lambda = \log_2 c, \quad v = \mathrm{O}\left(\log_2 \frac{n}{\varepsilon}\right); \tilde{n} = 2^v$$

Almost universal class of hash functions[6]

Definition. A family of mapping $H = \{h_i\}$ $\{0,1\}^n \to \{0,1\}^l$ is called δ-almost universal $(\delta - AU)$, if for any $x' \neq x : \Pr(h_i(x) = h_i(x')) \leq \delta$ under uniformly selected h_i from the set H.

In a particular case when $\delta = 2^{-l}$ we get a conventional family of universal hash functions.

Asymptotic behavior of parameters for some classes of AU hash functions has been considered in [6]. However constructive methods to design such hash functions are not sufficiently advanced. In application to authentication problem AU functions were used in [7].

Reducing AU to SE

Statement [8].

Let us consider any m, $\varepsilon > 0$ and $l \le m - 2l$. Then if $H = \{h_i : \{0,1\}^n \to \{0,1\}^l\}$ is AU for $\delta = 2^{-l}(1+\varepsilon^2)$, it results in the fact that *H* is SE.

Thus, if it is known how to design AU then it is known also how to design SE with some given parameters. But constructive methods to design AU are known not so much

Reducing of linear q-ary codes to SE

If $[T, k, (1 - 1/q - \delta)T]_q$ is some q-ary linear code with given parameters, then the exists $(1/\delta, \sqrt{q\delta/2})$ - extractor that can be presented as $Extr(\omega, x) = [C(\omega)]_{x=i}$, where $C(\omega)$ - code words corresponding to ω and $[\cdot]_{x=i}$ - is a random choice of i-th symbol.

Selection of SS parameters(see slide 16)

(m, m', t):

$H_\infty(W) = m$, $\tilde{H}_\infty(W \mid SS(W)) \ge m'$, $\rho(\omega, \omega') \le t$

$SS(W) = syn_c(\omega) = \omega H$,

where H is check matrix of some (n,n-r) linear code.
n is the length of the string ω, r is the number of check symbols .

Interconnection of the parameters n, r and t is due to Varshamov -Gilbert bound: r=nH(2t/n);
H(x)= − xlogx + (1-x)log(1-x).
How to determine the requirement to m'*?*
If the best method of statistical finding ω on SS(ω) is used, then the probability of success after $L = 2^s$ trails of ω is

$$P = 2^{s-m'}.$$

How to find m for BI ?
This is open problem . (Experimental testing with an estimation H(ω) and then an estimation of $H_\infty(\omega)$).

Open problems

1. Which of BI is preferentially?
2. How can we estimate $H_\infty(\omega)$, where ω is BI?
3. How can be established a secure level of $H_\infty(\omega/SS\ (\omega))$ for SS and ε for FE.
4. Constructive design of SE given its complexity.
5. Parameter optimization for SS and FE.
6. Parameter optimization for broadcast key distribution system based on FE technique.
7. Practical implementation of identification systems based on particular types of BI.
8. Design of SS and FE for Euclidian metrics on the plane.

References

1. **Zhang D.D. Automated Biometrics. Technologies and Systems. Wiley and Sons,2002.**
2. **Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data. *Lecture Notes in Computer Science* 3027, p. 523-540, *Springer-Verlag*, 2004.**
3. **Maurer U., Wolf S. Secret-key agreement over unauthenticated public channels - part III: Privacy amplification. IEEE Transactions on Information Theory, 2003, April, Vol. 49, No. 4, pp. 839 – 851.**
4. **V. Yakovlev, V. Korzhik, M. Bakaev, "Key Distribution protocols with the use of extractors based on noisy channels under the condition of active eavesdropper" Problems of Information Security (in Russian), N2, 2006, pg. 63-84.**
5. **Trevisan L. Construction of extractors using pseudo-random generator. Proceedings of the 31 annual ACM symposium on theory of computing, Atlanta, 1999, pp. 141 – 148.**
6. **Stinson D.R. Universal hashing and authentication codes. LNCS, v. 576.**
7. **Korjik V., Morales-Luna G. Hybrid authentication based on noisy channels . International Journal of information security, 2003, vol. 1, No. 4, pp.203 – 210.**
8. **Hasted J. et. al, Construction of pseudorandom generator from any one-way function. SIAM J. of Computing, 28(4), 1999, p. 1369 –1396.**
9. **Dodis Y. et. al, Robust Fuzzy Extractors and Authenticated Key Agreement from Close Secrets. Proc. of EUROCRYPT 2004.**
10. **Maurer U. Information-theoretically secure secret-key agreement by NOT authenticated public discussion . Advances in Cryptology – EUROCRYPT 97. Berlin, Germany: Springer-Verlag, 1997, vol. 1233, pp. 209 – 225.**
11. **V. Yakovlev, V. Korzhik, G. Morales-Luna, "Key Distribution Protocols Based on Noisy Channels in Presence of Active Adversary: Conventional and New Versions with Parameter Optimization", IEEE Trans. on IT, Special Issue on Information Security. (submitted, 2007)**
12. **Bernadette Dorizzi and Carmen Garcia-Mateo, "Multimedia Biometrics", Annals of telecommunications, vol.62,No.1,2, 2007.**

Combining crypto with biometrics in solution of user's identification problem.

Defects of SS- based approach:

1.Estimations of SS-security can fail sometimes.

Example. Consider iris as biometric information. It is binary string of the length 2048 bits with the mean intra-eye symbol error probability 0.127 [4] and the min entropy m=249 bits [5]. Then we get in line with Varshamov-Gilbert bound(SeeSl.28) that $r \geq 2048H(2\ t/h)$, where $t \sim 2048 \times 0.128 = 260$, and thus $r \geq H(0.254) \times 2048 = 0.81 \times 2048 = 1658$.It results in trivial inequality $H_\infty(W/SS(w)) \geq m-r = -1609$…?

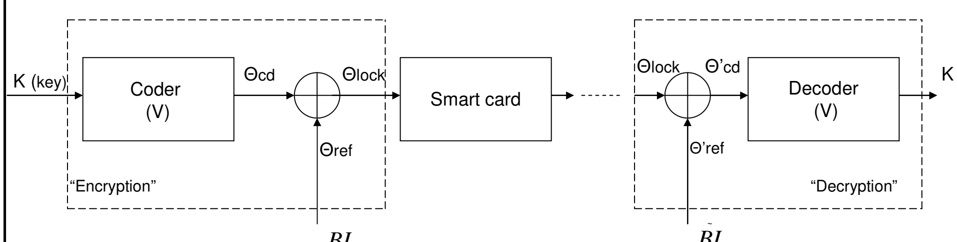
2.SS-based scheme fails completely whenever the original biometrics is stolen.

3.SS-based scheme is not key diversity one. It is inconvenient if user wishes to separate access key for his (her) bank account and to workplace computer.

How to remove these defects?

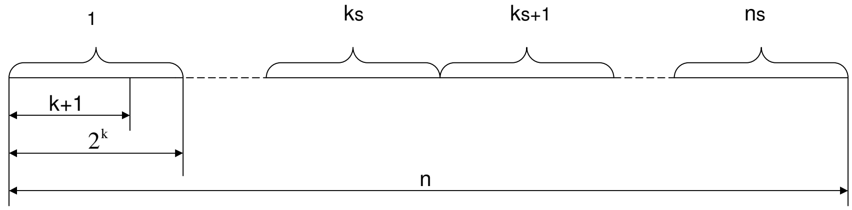
It is possible if user "encrypts" some truly random access-key by his(her) BI.
In this setting BI plays a role of encryption and decryption keys. But because BI varies in time it is necessary to reconcile this difference by error correcting code.

Basic scheme for iris BI



$\Theta`_{cd} = \Theta_{cd} \oplus \Theta_{ref} \oplus \Theta`_{ref} = \Theta_{cd} \oplus \mathbf{e}$, where **e** is error pattern between Θ_{ref} and Θ'_{ref}

The proposed coding scheme [4]



k- parameter of Hadamard code (HC),
k+1- the number of information bits of HC,
$\mathbf{2^k}$ - the length of HC,
k_s- the number of information bits of Reed-Solomon code (RSC),
n_s- the length of RSC,
$\mathbf{q = 2^{k+1}}$-the order of the field GF(q) connected with RSC,
$l=(k+1)k_s$- the total number of information bits of concatenated code (CC) that is equal to the length of the key K,
$\mathbf{n = 2^k n_s}$ -the length (in bits) of CC,
$\mathbf{t = 2^{k-2}}$ - error correction capability of HC,
$t_s = \frac{n_s - k_s + 1}{2}$ - error correction capability of RSC
HC- corrects bit errors
RSC- corrects both errors in blocks of HC and in addition bursts of errors

Proposed parameters
$k=6$, $k+1=7$, $\mathbf{2^k = 64}$, $k_s=20$, $n_s=32$, $t_s=6$, $t=16$, $n=2048$, $q=128$, $l=140$

Performance evaluation of the scheme above[6]

$$P_{FRR} \le \sum_{i=ts-t0+1}^{ns} \binom{ns}{i} pq^{i} \cdot (1-pq)^{ns-i},$$ (false rejection rate ~ the probability to reject an identification of valid person) (1)

where $$Pq \le \sum_{i=2^{k-2}+1}^{2^k} \binom{2^k}{i} \cdot p^{i} \cdot (1-p)^{2^k-i}$$

p- symbol error probability in the error pattern **e** for the same person
t_0- the amount of burst that corrects RSC.

$$P_{FAR} \le \sum_{i=0}^{ts} \binom{ns}{i} \cdot p'q^{i} \cdot (1-p'q)^{ns-i}$$, (false acceptance rate ~ the probability of positive identification of invalid person) (2)

where $$Pq' \le \sum_{i=2^{k-2}+1}^{2^k} \binom{2^k}{i} \cdot p'^{i} \cdot (1-p')^{2^k-i}$$

p`-the symbol error probability in the error pattern **e** for different persons.

The results of calculations P_{FRR}, P_{FAR} for different parameters CC and "channel parameters" p, p`, t_0.

We get for the proposed in [4] CC parameters:

$\mathbf{P_{FRR} \approx 2.2 \cdot 10^{-5}}$ $\mathbf{P_{FAR} \approx 7.35 \cdot 10^{-5}}$ if $t_0 \leq 6$, p=0.124, p`=0.3, l=140

It is possible to improve the efficiency of scheme if to select the following parameters: k=5, k+1=6, $\mathbf{2^k = 32}$, k_s=40, n_s=64, t_s=12, q=64

Then we get: $\mathbf{P_{FRR} \approx 1.9 \cdot 10^{-7}}$ $P_{FAR} \approx 1.9 \cdot 10^{-14}$ if $t_0 \leq 6$, and the key length l=240

Further improvement can be obtain by changing encoding / decoding procedures if we use HC in order to correct and detect errors, whereas RSC is used in order to correct both errors and erasures that decreases a complexity of decoding procedure.

Security of CC-based scheme.

Let us consider a situation where the token (smart card) is stolen while iris code remains unknown.
This means that Θ_{lock} is known by attacker while Θ_{ref} is unknown.
However a 2048-bit iris code has only 249 degrees of freedom [5].
Assume that attacker has perfect knowledge of the correlation within the subject's iris code.
Then the uncertainty of the iris code is only 249 bits.
The proposed coding scheme allows up to 27 percent of the bits to be wrong.
So the attacker is trying to find a 249 bit string within 67 bits Hamming distance of the key .
By the sphere-packing bound it requires to perform

$$T \geq \frac{2^{249}}{\sum_{i=0}^{67} \binom{249}{i}} \simeq \frac{2^{249}}{\binom{249}{67}} \approx 2^{44}$$ computations.

Attack by a compromization of the key

Let us assume that user has two different keys K and K`, which are “encrypted” by the same BI:

$\Theta_{lock} = f(K) \oplus \Theta_{ref}$ $\Theta`_{lock} = f(K) \oplus \Theta`_{ref}$

where f(…) is encoding functions,
θref, θ`ref are iris codes of the same person in different time moments.

Attacker knows: Θ_{lock}, $\Theta`_{lock}$ and K` and his other goal is to find K.
Then the attacker is able to find

$\Theta_{lock} \oplus \Theta`_{lock} = f(K) \oplus \Theta_{ref} \oplus f(K`) \oplus \Theta`_{ref} = f(K) \oplus f(K`) \oplus \Delta\Theta_{ref}$,

where $\Delta\Theta_{ref}=\Theta_{ref}\oplus\Theta`_{ref}$

Next attacker calculates $f(K)\oplus\Delta\Theta_{ref}=\Theta_{lock}\oplus\Theta`_{lock}\oplus f(K`)$
Since the error pattern $\Delta\Theta_{ref}$ can be corrected by CC attacker is able to find the second key K. (This attack was not mentioned in [4]).

Nonlinear encryption of the secret identification key by BI (Fuzzy Vault (FV) scheme [1])

Encryption

1. Sekret key $K \to P_K(z)=\sum_{i=1}^{k} k_i \; z^{i-1}$, where $K \to (k_1, k_2, \dots, k_k)$, $k_i \in GF(q)$
2. $BI \to (x_1, x_2, \dots, x_t)$ $x_i \in GF(q)$, $t > k$
3. $y_i = P_K(x_i)$, $i = 1,2,\dots,t$, $y_i \in GF(q)$
(y_i - code word of q-ary (t,k) Reed Solomon Code (RSC))
4. Random generation $R=\{\tilde{x}_i, \tilde{y}_i\}, i=1,2,\dots,r-t$, $r>t$, where $\tilde{x}_i \neq x_j, \tilde{y}_i \neq y_j, \forall i,j$
5. $FV \to S=\{S_x, S_y\}=\phi(T,R)$
where $T=\{x_i, y_i\}_{i=1}^{t}$, $R=\{\tilde{x}_i, \tilde{y}_i\}_{i=1}^{r-t}$, $\phi(..,..)$ - random shuffling of pairs

Decryption

1. To get $FV \to S$
2. $BI \to T'_x=\{x'_i\}_{i=1}^{t}$

3. $P_x = \{ s \in S_x \; : \; s = \underset{s \in S_x}{Argmin} \, \rho(s, z), \; z \in T'_x \}$,

where $\rho(s, z)$ is some metric (distance between s and z)

4. $P_x \rightarrow (P_x, P_y)$, where $P_y = \{ y \in S_y \; : \; x \in P_x \}$

5. $(P_x, P_y) \rightarrow \tilde{P}_K(z) \rightarrow \tilde{K}$ (correction of errors and erasures by RSC)

Modification of the decryption step (3)

3'. $P'_x = \{ s \in S_x \; : \; s = \underset{s \in S_x}{Argmin} \, \rho(s, z), \; \rho(s, z) \leq t_0 \; z \in T'_x \}$

Particular cases of metrics ρ(s,z)

1. Hamming metric:

$$\rho(S, Z) = \rho_H(S, Z) = \{ \; l : S_l \neq Z_l \;\; S = \{ S = l \}_1^m, \; \{ Z = l \}_1^m \; \}$$

2. Euclidean metric:

$$\rho(S, Z) = \sum_{i=1}^{m} (S_l - Z_l)^2$$

3. Set distance metric:

$$P_x = S_x \cap T'_x$$

Error correction capability of RSC

$d = t - k + 1$ (minimal code distance)

$\delta = \frac{t-k}{2}$ (the maximum number of errors that can be corrected for sure)

If RSC corrects δ' erasures it is still able to correct in addition $\delta_0 = \frac{t-k-\delta'}{2}$ errors for sure

Example

Let us assume that a set of favorable movies is BI. $L = 10^4$ (list of movies), $q = 2^{14}$, $t = 22$, $k = 14$. The total number of secret keys K is $(2^{14})^{14} = 2^{196}$, $d = 22 - 14 + 1 = 9$, $\delta = 4$ either 4 erasures and 2 errors. This mean that another person can get access to E-mail address some person which distribute his (or her) FV on Internet if and only if they share at least 18 favorable movies.

Security of FV scheme

Goal of an attacker

To find secret key K given FV and the full knowledge of its design.

Brute force attack

Select randomly k elements from the set S_x given $FV=(S_x, S_y)$. If the chosen set $A \subseteq T_x$, where $T_x = \{x_i\}_{i=1}^{t}$, then it results in a revealing of the secret K, because polynomial can be correctly interpolated by its k values.

The probability that such attack be successful is:

$$P_{sa} = \frac{\binom{t}{k}}{\binom{r}{k}} \Rightarrow P_{sa}^{-1} = \frac{\binom{r}{k}}{\binom{t}{k}} \sim \left(\frac{r}{t}\right)^k \le 1.1\left(\frac{r}{t}\right)^k, \text{for } r > t > 5 \quad [3]$$

How to verify that the key obtained after attack is correct?

1. Apply this key to decrypt some address or to be identified (but this way is possible not always).

2. If the recovered polynomial $\tilde{P}_K(z)$ is correct that is $\tilde{P}_K(z)=P_K(z)$ then: $P_k(x_i)=y_i$ for at least t-k values $x \in T_x$. Otherwise it is true only with the probability q^{-1}. This fact allows to verify if the key chosen by attacker is correct or not. The attacker can recover the secret key K in $L \sim 8 \cdot k \log^2(k) \cdot \left(\frac{r}{t}\right)^k$ operations [3]

Examples of the use FV for different BI

1. Iris code (Binary strings of the length 2048)

Let us select: $q=2^{16}, t=2^{7}, k=10$ (this provides 160 bit key length). Then RSC has $\delta = \frac{128-10}{2} = 59$ error capability for sure. If we let the probability of bit error for intra-eye with mask is of about 0.03 [4] then the q-ary symbol probability will be

$$1-(1-0.03)^{16} \approx 0.4$$

The averaged number of q-ary errors on the block of RSC is $128 \cdot 0.4 \approx 51$

The probability of incorrect decoding by RSC is:

$$P_{ed} \approx \sum_{i=60}^{128} \binom{128}{i} (0.4)^i (0.6)^{128-i} \le \left(\frac{0.6}{0.4} \cdot \frac{60}{128-60} \right)^{-60} \cdot \left(0.6 \left(1 + \frac{60}{128-60} \right) \right)^{128} \approx 0.289$$

In order to provide security of FV we have to select $L \approx 2^{60}$ for $k = 10$ (see previous slide). Thus we have to select $\binom{r}{t} \approx 2^5$ and hence $r = 2^5 \cdot 2^7 = 2^{12}$. This means that total size of FV memory is about $4 \cdot 2^{12} = 2^{14}$ bytes that can be too much for ordinary smart card. Thus application of FV with iris code is limited.

2. Fingerprinting

Typical example [2]:

$q = 251^2, r = 1000, t = 40, k = 80$ It gives $L \sim 2^{58}$

But in [3] has been shown that there exists an improved brute force attack that is able to break FV with chosen above parameters and proposed some modifications of FV schemes which are resistant against such attack.

3. Sharing of tastes in movies

This application of BI encryption is provided well by FV technique [1].

(See example in previous slides).

Conclusion and open problems

1.Combining crypto with biometrics provides additional security level for user identification if the original biometrics are stolen.

2.Combinig scheme (CS) may solve key diversity problem (the use of different access keys for different identification points) only if precautions are taken in addition to CS . Generally speaking , it is still open problem.

3.A choice of the CS type depends on the types of biometrics. So for iris BI seems to be better to use CS based on concatenated codes whereas FV is better both for fingerprinting BI and for sharing of tastes BI .

4.Security of CS can be considered only as partly solved problem because not all attacks on them have been investigated in details.

5.Nevertheless CS are looking as very perspective methods and they have promising practical applications and offer interesting open problems for further theoretical investigation.

References

1. A. Juels and M. Sudan, “A fuzzy vault scheme”, Proc. of the IEEE International Symposium on Information Theory, 2002, p. 408.

2. C. Clancy, N. Kiyavash and D. Lin, "Secure Smartcard - Based Fingerprint Authentication”, ACM Workshop on biometric methods and applications, WBMA'03.

3. P. Mihailescu, “The Fuzzy Vault For Fingerprints is Vulnerable to Brute Force Attack”, http://arxiv.org/abs/0708.2974.

4. F. Hao et. al, “Combining Crypto with Biometrics Effectively”, IEEE Trans. on Computers, vol. 55, №9 , 2006, p. 1081-1088.

5. J. Daugman, “The Importance of Being Random: Statistical Principles of Iris Recognition”, Pattern Recognition, 36, 2003, p. 279 - 291.

6. V. Korzhik, E. Nikolaeva, “Identification Based on Biometric Information”, VIII International Simposium on Intellectual Systems, 2008, (submittal).