Finger Print Identity & Security

Daniel Novák

5.1. 2016, Prague

Acknowledgments: Michal Huptych, Vaclav Chudacek, Xavier Palathingal, Andrzej Drygajlo, Handbook of Fingerprint Recognition

laboratory Gerstner

Outline

- Identity

Only once during the existence of our solar system will two human beings be born with similar finger markings (Harper, 1910)

Two like fingerprints would be found only once every 10^48 years

Dabert *court – Biometrics must meet 5 conditions*
Handwriting is not accepted

- Security
  - **Easiest way: bribe system admin☺**

laboratory Gerstner

What does biometric individuality mean?

- Given a biometric sample, determine the probability of finding an arbitrary biometric sample from the target population sufficiently similar to it. (Pankanti *et al.*)

- In other words, what are the theoretical lower bounds on the FAR and FRR, often called the "*intrinsic error rates*".

- FP identification is based on:
  - (i) persistance-the basic charateristcs does not change with time
  - **(ii) FP is UNIQUE to an individual ->** not scientifically established ->the validity of FP is now being challenged in several court cases!!!!!

How to estimate?

- REPRESENTATION
  - Previous lectures
- SIMILARITY METRICS
  - Empirical approach
  - Model approach

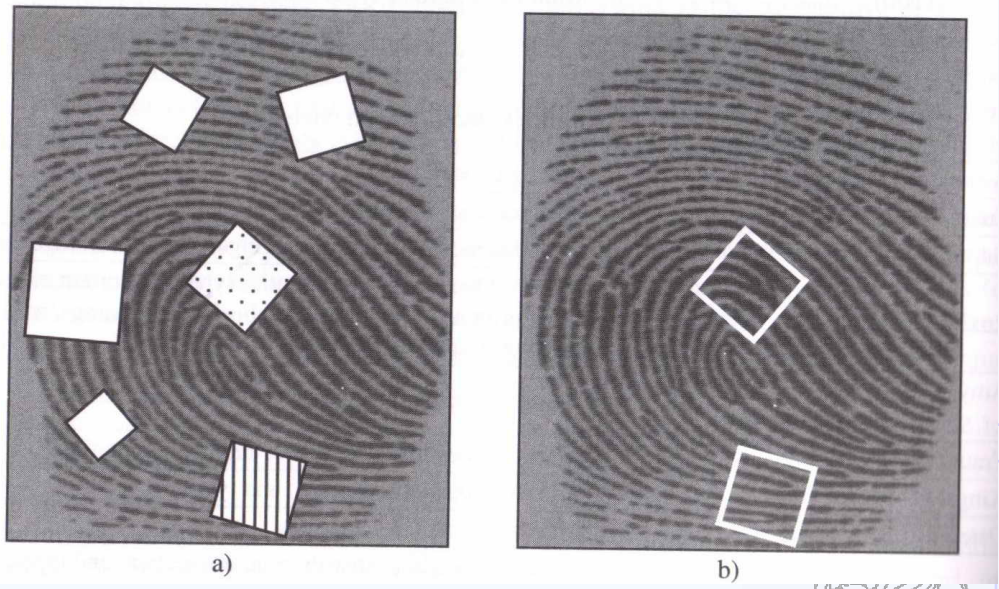
- Empirical approach -> 200 millions in FBI register
- 1270 years to estimate with speed of 1 million matches per second (200x10^6x200x10^6/10^6x60x60x24x365=1270!)

laboratory Gerstner

Finger print configuration

$$P(\text{Fingerprint Configuration}) = \frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{2}\right)^{24} = 1.45 \times 10^{-11}$$

- ½ probability of reconstruction
- 24 boxes with 6 ridges can cover FP
- 1/16 – probability of occurrence of a specific fingerprint (such as arch, left loop)
- Occurrence of correct number of ridges entering and existing each of 24 boxes



$$P(\text{Fingerprint Configuration}) = p^N.$$

Different models

Table 2: Comparison of probability of a particular fingerprint configuration using different models. For a fair comparison, we do not distinguish between minutiae types. By assuming that an average size fingerprint has 24 regions ($R = 24$) as defined by Galton, 72 regions ($M = 72$) as defined by Osterburg et al., and has 36 minutiae on an average ($N = 36$), we compute the probability of observing a given fingerprint configuration in the third column of the table. The probability of observing a fingerprint configuration with $N = 12$, and equivalently, $R = 8$ and $M = 24$, is given in braces in the third column. Note that all

Author	P(Fingerprint Configuration)	N=36,R=24,M=72 (N=12,R=8,M=24)
Galton (1892)	$\frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{2}\right)^R$	1.45×10^{-11} (9.54×10^{-7})
Pearson (1930)	$\frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{36}\right)^R$	1.09×10^{-41} (8.65×10^{-17})
Henry (1900)	$\left(\frac{1}{4}\right)^{N+2}$	1.32×10^{-23} (3.72×10^{-9})
Balthazard (1911)	$\left(\frac{1}{4}\right)^N$	2.12×10^{-22} (5.96×10^{-8})
Bose (1917)	$\left(\frac{1}{4}\right)^N$	2.12×10^{-22} (5.96×10^{-8})
Wentworth & Wilder (1918)	$\left(\frac{1}{50}\right)^N$	6.87×10^{-62} (4.10×10^{-21})
Cummins & Midlo (1943)	$\frac{1}{31} \times \left(\frac{1}{50}\right)^N$	2.22×10^{-63} (1.32×10^{-22})
Gupta (1968)	$\frac{1}{10} \times \frac{1}{10} \times \left(\frac{1}{10}\right)^N$	1.00×10^{-38} (1.00×10^{-14})
Roxburgh (1933)	$\frac{1}{1000} \times \left(\frac{1.5}{10 \times 2.412}\right)^N$	3.75×10^{-47} (3.35×10^{-18})
Trauring (1963)	$(0.1944)^N$	2.47×10^{-26} (2.91×10^{-9})
Osterburg et al. (1980)	$(0.766)^{M-N}(0.234)^N$	1.33×10^{-27} (1.10×10^{-9})
Stoney (1985)	$\frac{N}{5} \times 0.6 \times (0.5 \times 10^{-3})^{N-1}$	1.2×10^{-80} (3.5×10^{-26})

- ¼ - four types of equally likely minutiae events
  - Bifurcation to the left
  - Bifurcation to the right
  - Ending to the left
  - Ending to the right

Jain model-matching

- The probability of false correspondence between two fingerprints belonging to different fingers
- 7 assumption including that a reasonable alignment has been established
- The probability that FP with 36 minutiaes will share 12 minutiaes with another arbitrarily chosen FP with 36 minutias is 6.1x10^-8
- REPRESENTATION

$$T = \{\{x_1, y_1, \theta_1\}, \{x_2, y_2, \theta_2\}, ..., \{x_m, y_m, \theta_m\}\},$$
$$I = \{\{x'_1, y'_1, \theta'_1\}, \{x'_2, y'_2, \theta'_2\}, ..., \{x'_n, y'_n, \theta'_n\}\}.$$

- Two FPs match if

$$\sqrt{(x'_i - x_j)^2 + (y'_i - y_j)^2} \leq r_0, \quad \text{and}$$
$$\min\left(|\theta'_i - \theta_j|, 360 - |\theta'_i - \theta_j|\right) \leq \theta_0,$$

Jain model-prob. match

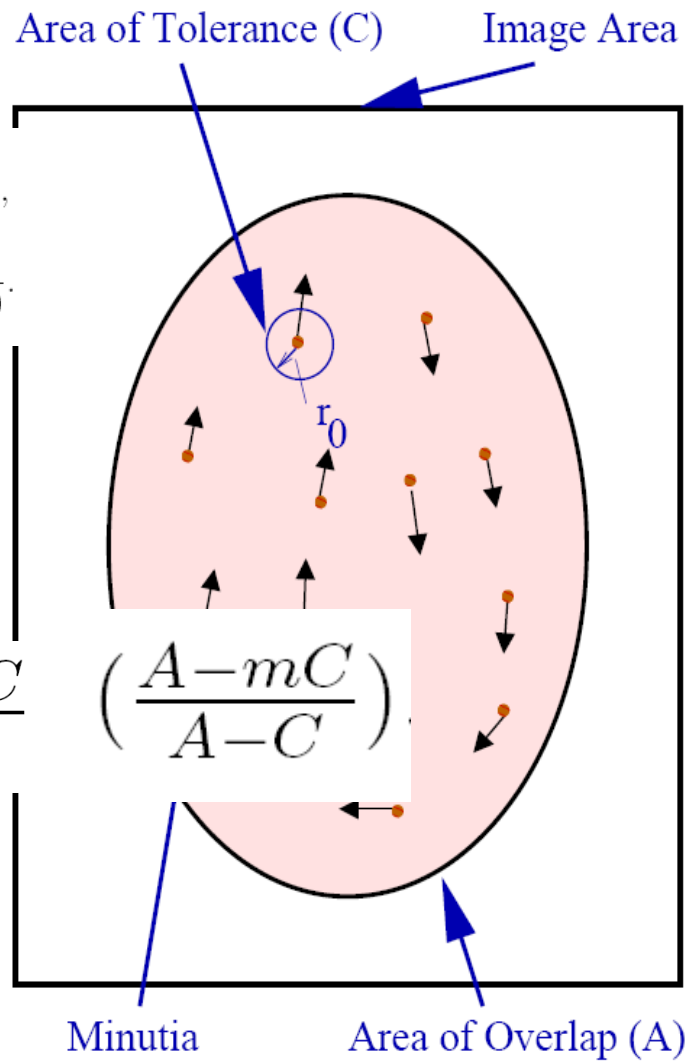
- The prob of matching minutias

$$P\left(\sqrt{(x_i'-x_j)^2+(y_i'-y_j)^2}\le r_0\right)=\frac{\text{area of tolerance}}{\text{total area of overlap}}=\frac{\pi r_0^2}{A}=\frac{C}{A},$$

$$P\left(\min\left(|\theta_i'-\theta_j|,360-|\theta_i'-\theta_j|\right)\le\theta_0\right)=\frac{\text{angle of tolerance}}{\text{total angle}}=\frac{2\theta_0}{360}.$$

- One input minutia matches any of m template minutias $\frac{mC}{A}$
- Two input minutiaes, only first matches, second FALSE accep $\frac{mC}{A}\left(\frac{A-mC}{A-C}\right)$
- First or second can correspond

$$2\times\frac{mC}{A}\times\frac{A-mC}{A-C}$$



Jain model-ρ match *m*

- ONE input (*n*) minutias matches one of the *m* minutiaes is

$$p(A,C,m,n) = \binom{n}{1} \left(\frac{mC}{A}\right) \left(\frac{A-mC}{A-C}\right)$$

- *r*o input minutiaes correnspond, n-ro doesn not correspond any match

$$p(A,C,m,n,\rho) = \binom{n}{\rho} \underbrace{\left(\frac{mC}{A}\right)\left(\frac{(m-1)C}{A-C}\right)\ldots\left(\frac{(m-\rho-1)C}{A-(\rho-1)C}\right)}_{\rho \text{ terms}} \times$$
$$\underbrace{\left(\frac{A-mC}{A-\rho C}\right)\left(\frac{A-(m-1)C}{A-(\rho+1)C}\right)\ldots\left(\frac{(A-(m-(n-\rho+1))C}{A-(n-1)C}\right)}_{n-\rho \text{ terms}}.$$

Jain model-direction

- After rearranging, where $M = \frac{A}{C}$

$$p(M, m, n, \rho) = \frac{\binom{m}{\rho}\binom{M-m}{n-\rho}}{\binom{M}{n}}.$$

•q minutiae among ρ have similar directions

$$(q \leq \rho)$$

$$P\left(\min\left(\left|\theta_i' - \theta_j\right|, 360 - \left|\theta_i' - \theta_j\right|\right) \leq \theta_0\right) = l$$

- Including direction l probability of two position – matched minutiae having a similar direction and *1-l i*s the probability of two-matched minutiae taking different directions,

$$p(M, m, n, q) = \sum_{\rho=q}^{\min(m,n)} \left(\frac{\binom{m}{\rho}\binom{M-m}{n-\rho}}{\binom{M}{n}} \times \binom{\rho}{q} (l)^q (1-l)^{\rho-q} \right).$$

Parameters to est

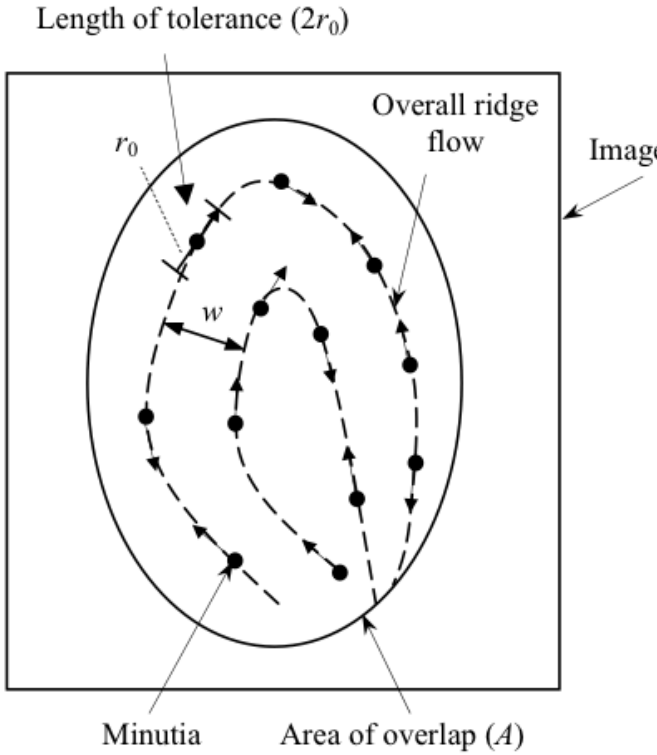
- Until now: minutiae locations are uniformly distributed within the *entire* area
- Ridges occupy *A/2* area, minutiae lie only on ridges, i.e., along a curve of length *A/w, w* ridge period

$$M = \frac{A/w}{2r_0}$$

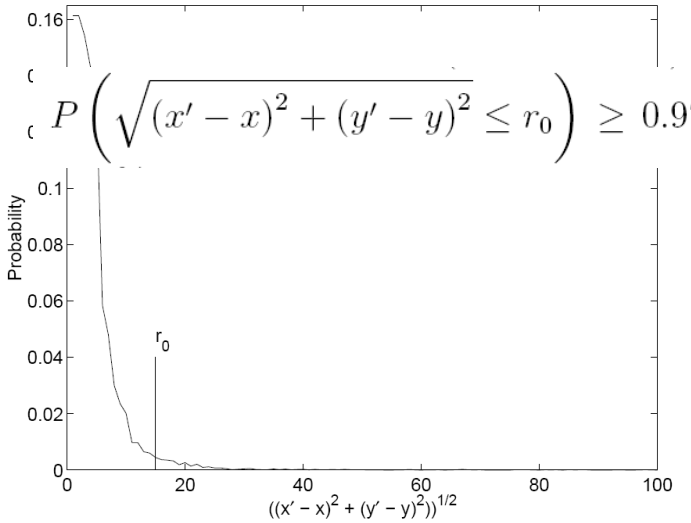
r_0, θ_0, l, and w

A, m, and n

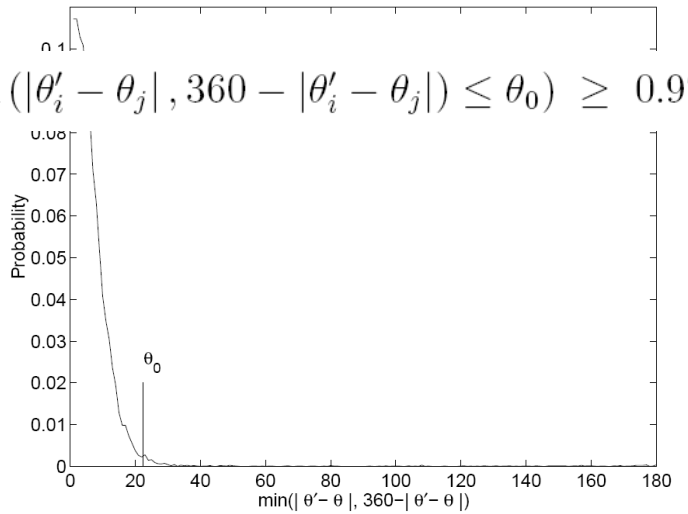
- Parameters from a given sensor resolution:
- 500 dpi (r0=15 pixels,22.5°), w=0.436mm



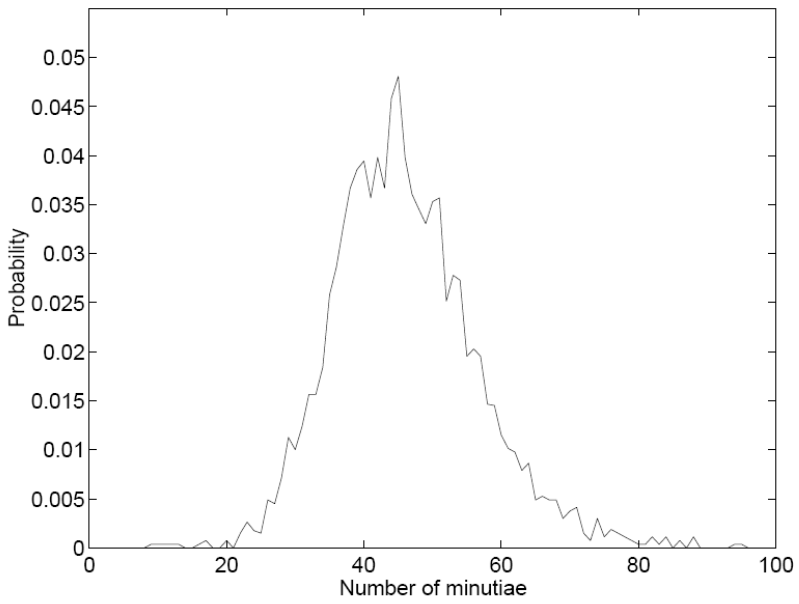
$$P\left(\sqrt{(x'-x)^2+(y'-y)^2} \le r_0\right) \ge 0.975$$

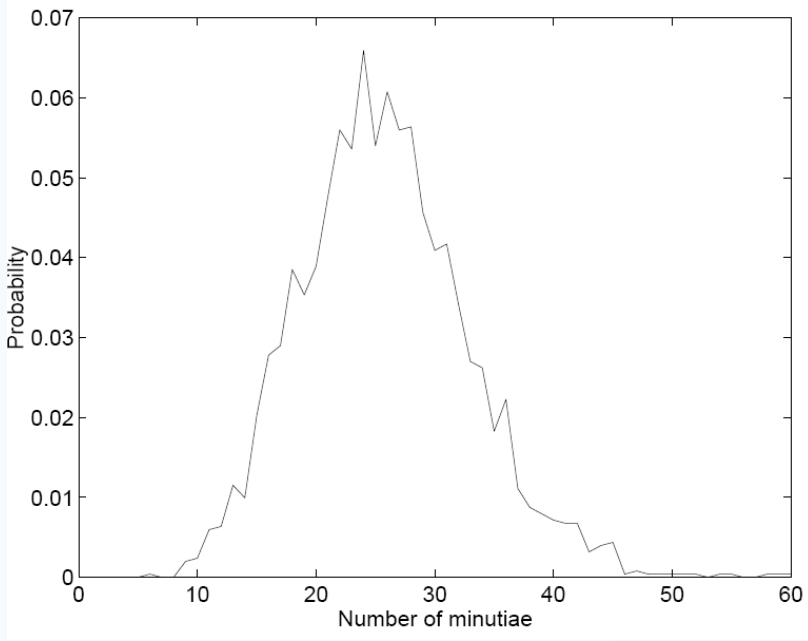


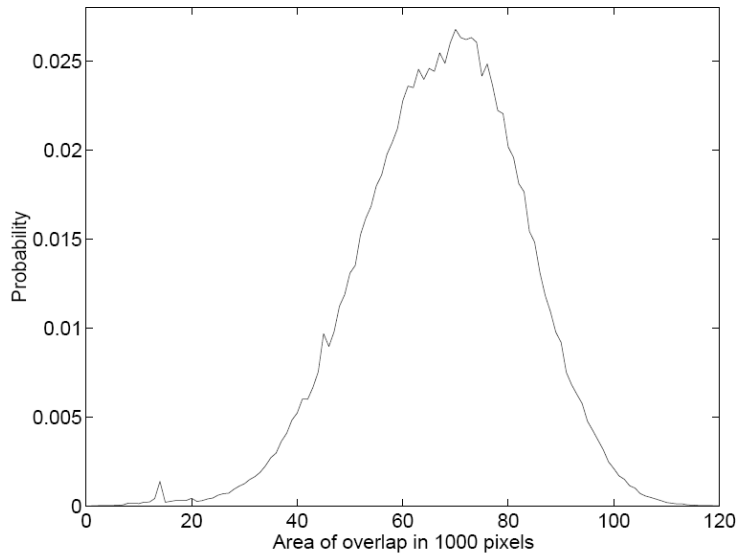
$$P\left(\min\left(|\theta_i' - \theta_j|, 360 - |\theta_i' - \theta_j|\right) \le \theta_0\right) \ge 0.975.$$

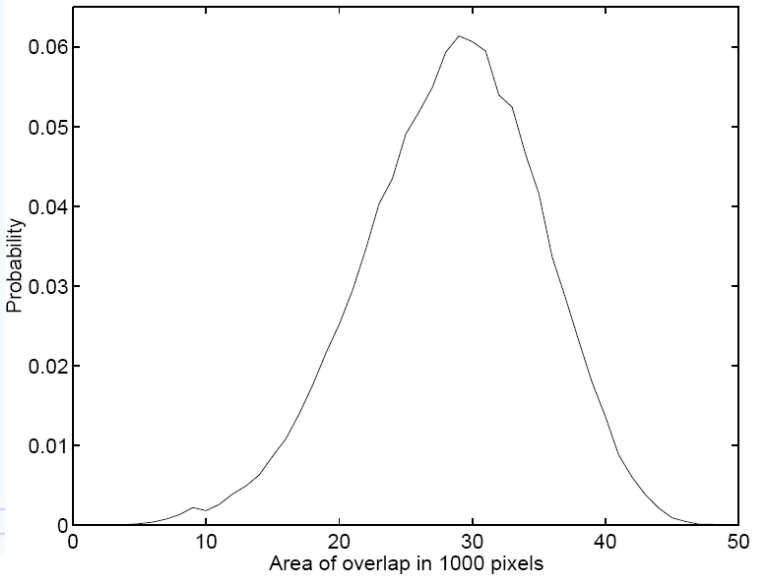


A,n,m


Probability
Number of minutiae
0.05
0.045
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
20
40
60
80
100


Probability
Number of minutiae
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50
60


Probability
Area of overlap in 1000 pixels
0.025
0.02
0.015
0.01
0.005
0
0
20
40
60
80
100
120


Probability
Area of overlap in 1000 pixels
0.06
0.05
0.04
0.03
0.02
0.01
0
0
10
20
30
40
50

Jain model - results

M, m, n, q	P(Fingerprint Correspondence)
104, 26, 26, 26	5.27×10^{-40}
104, 26, 26, 12	3.87×10^{-9}
176, 36, 36, 36	5.47×10^{-59}
176, 36, 36, 12	6.10×10^{-8}
248, 46, 46, 46	1.33×10^{-77}
248, 46, 46, 12	5.86×10^{-7}
70, 12, 12, 12	1.22×10^{-20}

- WEAK password (birthday, spouse's name), guessing by brute force
- 1/(26+26+10)^6=1.76x10^-11

Jain model – 12 guidelines

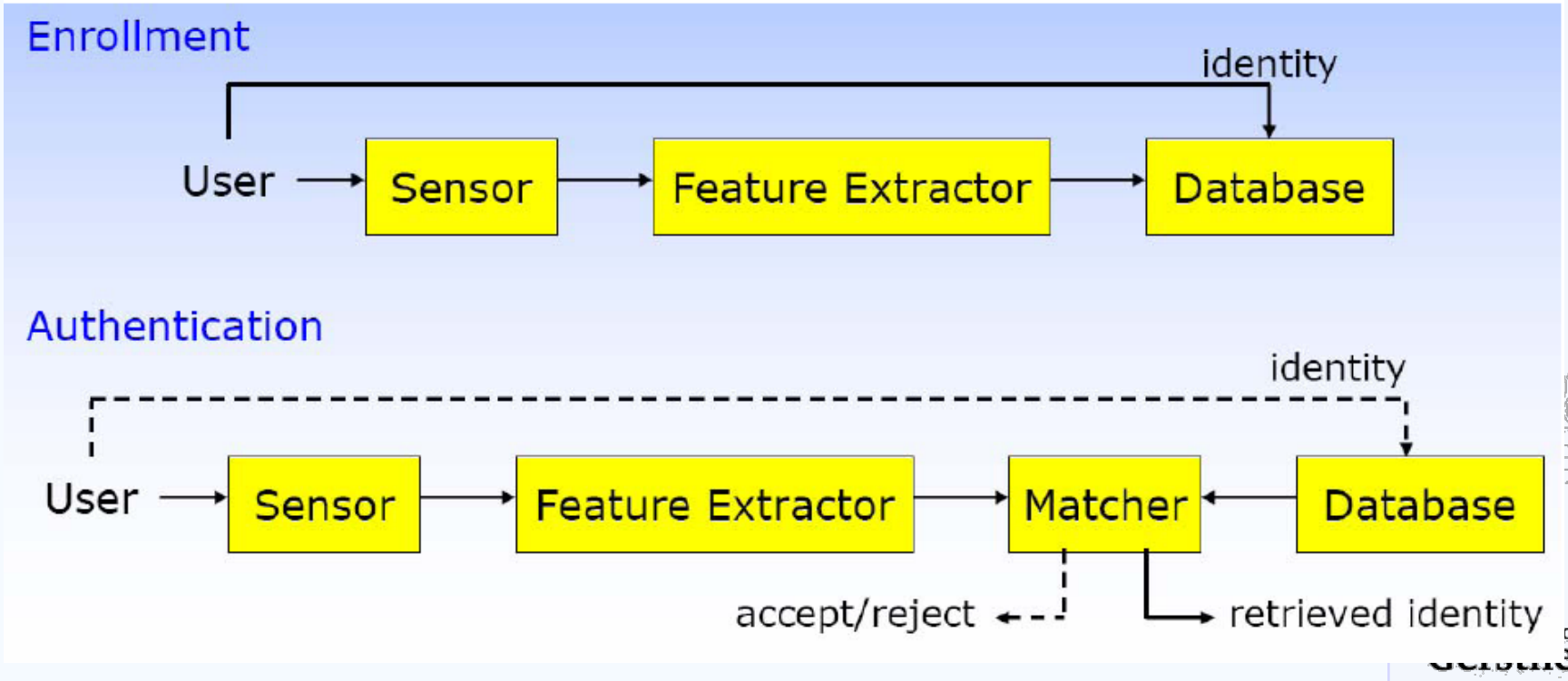
- 12 point guideline as sufficient evidence in many courts of law

q / n	8	9	10	11	**12**
12	6.19×10^{-10}	4.88×10^{-12}	1.96×10^{-14}	3.21×10^{-17}	$\mathbf{1.22 \times 10^{-20}}$
13	1.58×10^{-9}	1.56×10^{-11}	8.42×10^{-14}	2.08×10^{-16}	1.58×10^{-19}
14	3.62×10^{-9}	4.32×10^{-11}	2.92×10^{-13}	9.66×10^{-16}	1.11×10^{-18}
15	7.63×10^{-9}	1.06×10^{-10}	8.68×10^{-13}	3.60×10^{-15}	5.53×10^{-18}
16	1.50×10^{-8}	2.40×10^{-10}	2.30×10^{-12}	1.45×10^{-14}	2.21×10^{-17}

Table 4: The adverse effects of the fingerprint expert misjudgments in using the *12-point guideline*. The source of error could be in underestimating the number of actual minutiae in the latent print (n) or overestimating the number of matched minutiae (q). The value of m is 12 for all the entries in this table. The entry $(n = 12, q = 12)$ represents probability of a false correspondence when the 12-point guideline is correctly applied by a fingerprint examiner. Except for $(n = 12, q = 12)$ entry, all other entries represent incorrect judgements by the fingerprint expert to arrive at a decision that exactly 12 minutiae in the latent print matched 12 corresponding minutiae in the template print. For instance, the entry $(n = 14, q = 8)$ in the table represents an estimate of probability of a false correspondence due to two misjudgements by the examiner: Firstly, the fingerprint examiner detected 12 minutiae in the latent print while there were in fact 14 minutiae in the latent print, i.e., the examiner overlooked 2 latent print minutiae; Further, while he associated all the 12 minutiae he detected in the latent print to the 12 minutiae in the template print, only 8 of those correspondences were indeed genuine correspondences (4 incorrect minutiae match judgments).

Enrollment & Authentication

- Enrollment: User's biometric data is captured and a salient feature set is extracted; these features are associated with the user identity and stored as a template in a database
- Authentication: User's biometric data is captured and the extracted feature set is compared with either (i) all the templates in the database (identification), or (ii) the templates associated with a claimed identity (verification)



Why security?

- The number of installed biometric systems in both commercial and government sectors is increasing
- The size of the population that uses these systems is increasing (tens of millions in the US VISIT program)
- New application areas are emerging (visa, border control, e-commerce, health care records, entertainment …)

- Hence, the potential damage resulting from security breaches in biometric systems can be enormous

- Security analysis of biometric systems is critical

Threats

The Modern Burglar

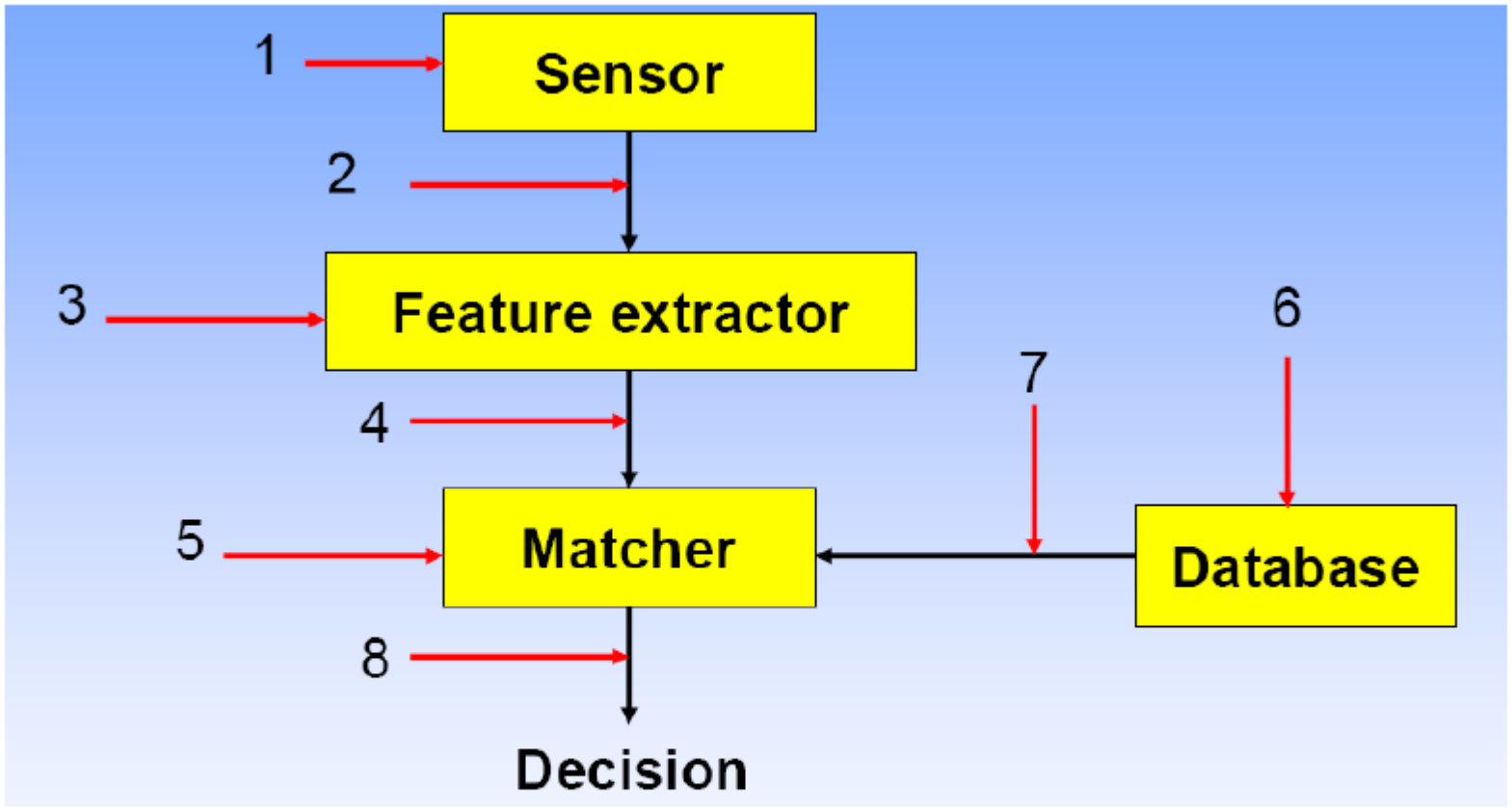
Types of threats

- Circumvention: An attacker gains access to the system protected by biometric authentication
  - Privacy attack: Attacker accesses the data that she was not authorized (e.g., accessing the medical records of another user)
  - Subversive attack: Attacker manipulates the system (e.g., submitting bogus insurance claims)
- Repudiation: An attacker denies accessing the system
  - A bank clerk modifies the financial records and later claims that her biometric data was stolen and denies that she is responsible
- Contamination (covert acquisition): An attacker illegally obtains biometric data of genuine users and uses it to access the system
  - Lifting a latent fingerprint and constructing a synthetic finger

Types of threats

- Collusion: A user with wide super user privileges (e.g., system administrator) illegally modifies the system
- Coercion: An attacker forces a legitimate user to access the system (e.g., using a fingerprint to access ATM at a gunpoint)
- Denial of Service (DoS): An attacker corrupts the biometric system so that legitimate users cannot use it
  - A server that processes access requests can be bombarded with many bogus access requests, to the point where the server's computational resources can not handle valid requests any more.

Threats locations



Points of attack for a generic biometric system

Threats locations

- Attack 1: A fake biometric (e.g., an artificial finger) is presented at the sensor
- Attack 2: Illegally intercepted data is resubmitted (replay)
- Attack 3: Feature detector is replaced by a Trojan horse program
  - It produces feature sets chosen by the attacker
- Attack 4: Legitimate features are replaced with a synthetic feature set
- Attack 5: Matcher is replaced by a Trojan horse program
  - It produces scores chosen by the attacker
- Attack 6: Templates in the database are modified, removed, or new templates are added
- Attack 7: The transferred template information is altered in the communication channel
- Attack 8: The matching result (e.g., accept/reject) is overridden

Fake from minutia data

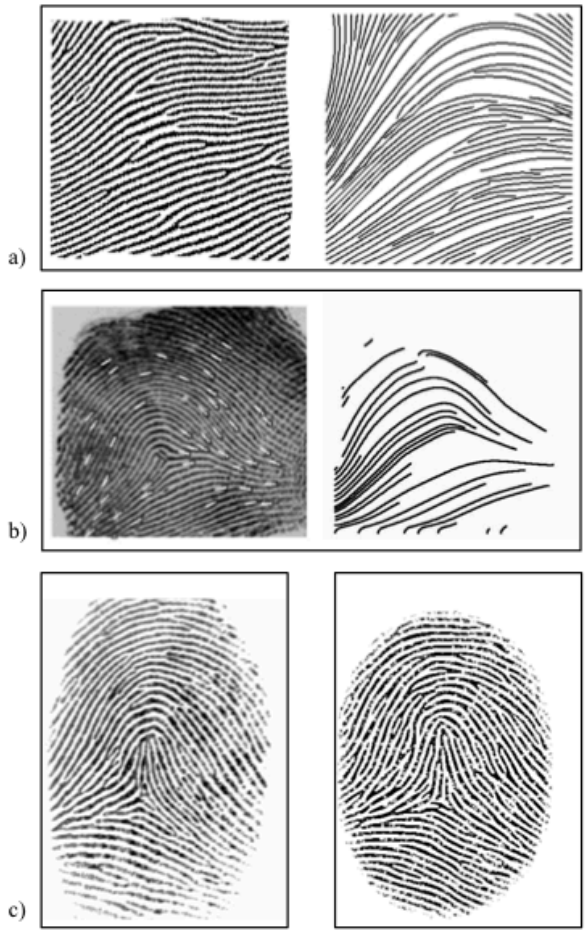


Figure 9.3. Three examples of fingerprint image reversal: the images shown on the right were reversed from the minutiae of the corresponding fingerprint images on the left: a) Hill (2001) method; although the two images do not visually look the same, they are similar enough for an automatic fingerprint recognition system to result in a match decision; b) Ross, Shah, and Jain (2005, 2007) method; c) Cappelli et al. (2007) method produces images that are visually quite realistic; even if these realistic reversals can easily fool an automatic fingerprint recognition system, they cannot fool a human expert. © IEEE. Images courtesy of C. J. Hill.

Attack 1 Example

Attack 1: Synthetic Biometric Submission

- No detailed system knowledge or access privileges is necessary
- Digital protection mechanisms (e.g., encryption) are not applicable

Putte, Keuning 2000:

- 6 fingerprint verification systems attacked
- 5 out of 6 accepted the dummy finger in the first attempt

Dummy finger created with cooperation of the user in a few hours with liquid silicon rubber

Dummy finger created from a lifted impression of the finger without cooperation of the user in eight hours with silicon cement

Attack 1 example

Matsumoto et al. 2002:

- 11 fingerprint verification systems attacked with artificial gelatin fingerprints
- Gelatin fingers accepted with a probability of 67-100%

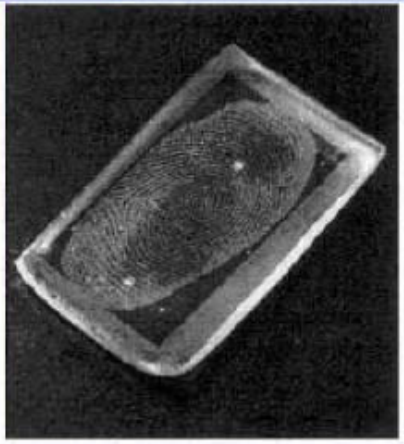
live

gelatin

With cooperation (finger pressed to plastic mold)

mold

gelatin



Without cooperation (residual fingerprint lifted from a glass)

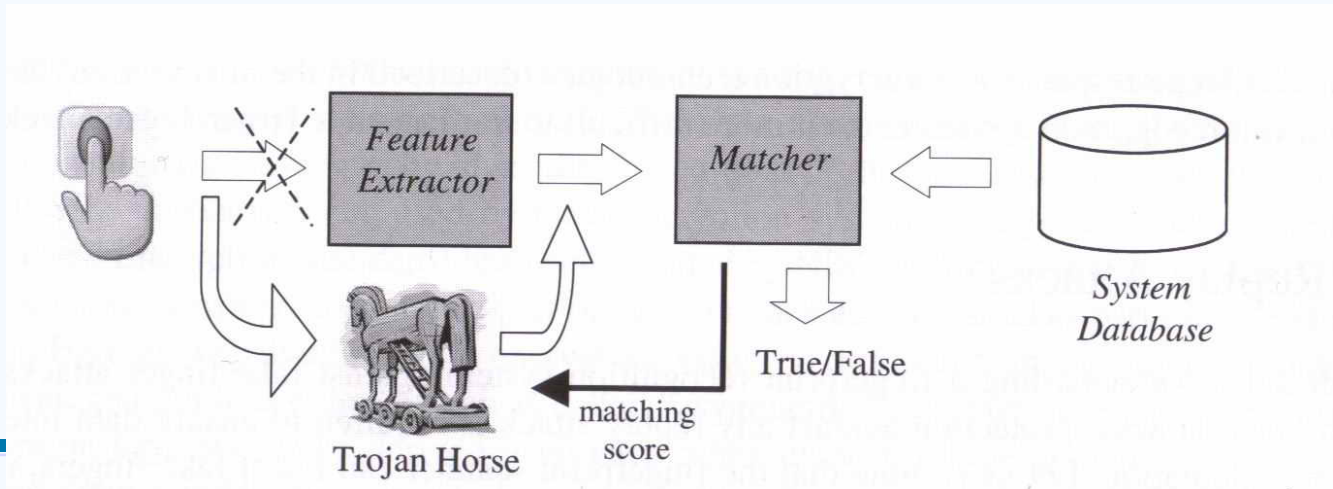
Finger vitality detection

- Pattern classifier to discriminate between human and synthetic epidermis
- Vital signs
  - Temperature: at 20C epidermis higher by 8 to 10 degrees, fake finger 2 degrees less, cold coca-cola can
  - Conductivity: greatly varies, water or saliva is added to fake
  - Optical sensors: meassure absorption, reflection, scattering, refraction, gelation has similar optical properties to a live finger
  - Ultrasonic sensors: detec layer under epidermis, silicon rubber layer + silicon rubber finger
  - Increase resolution, detect sweet pores
  - Blood pressure, ECG meassurement
  - VIDEO

laboratory Gerstner

Attack 5 – Trojan horse

- Sensor emulator, feature extraction, matcher, system database
- Digital signature
- Trust authorities – standard electronic commerce systems
- PIN example, 4 digits, 10^4 combinations
  - A) know password: easy
  - B) brute force attack
- FP example (API downloadable??? E.g.BioAPI: http://www.bioapi.org/)
  - FP representaion MUST be know: type of features, digital representation, quantization, spatial reference, ordering, etc.
  - Randomly generate FP representation, better synthetic generator, *apriori knowledge from latent fingerprints*
  - Gray-scale: use synthetic generator
- **Hill climbing**, match feedback available, iteratively details changing after positive feedback

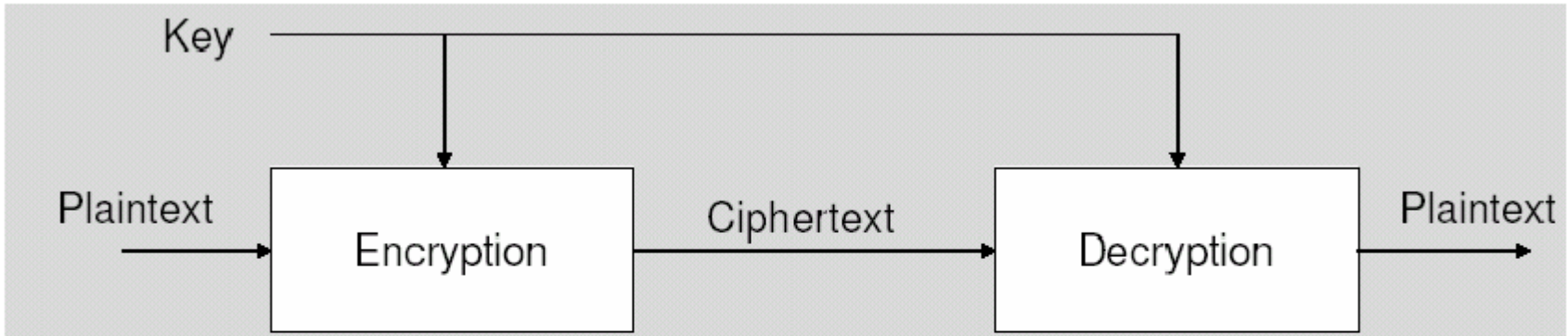


Cryptography

- In biometrics systems the integrity of data transmission must be secure all the way from the sensor to the application. This is typically achieved by cryptographic methods
- The enhancement of security level in biometrics-based systems can be done in two ways, not necessary mutually exclusive:
  - Keys/biometrics securing biometrics/keys: use of encryption keys to protect biometric information (for authentication purposes) or use of biometric mechanisms to secure the privacy of encryption keys and access to data

Symetrické šifrování

- Šifrovací algoritmus, který používá k šifrování i dešifrování jediný klíč
- Výhodou symetrických šifer je jejich nízká výpočetní náročnost
- Velkou nevýhodou je nutnost sdílení tajného klíče
- DES-56 bitovy klic
- AES (Advanced Encryption Standard) – standard by NIST



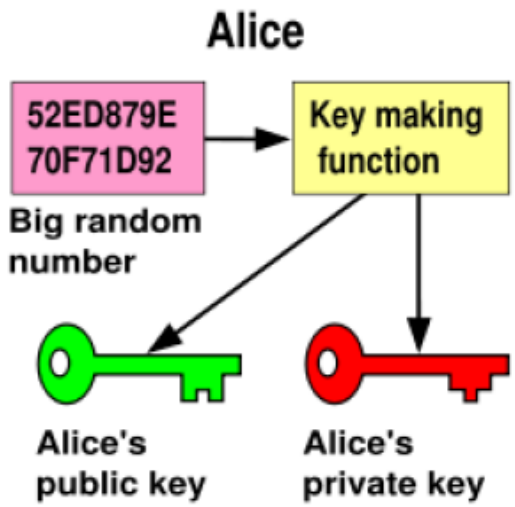
- Uses a single key
- Examples
  - DES (Data Encryption Standard)

Asymetrické šifrování

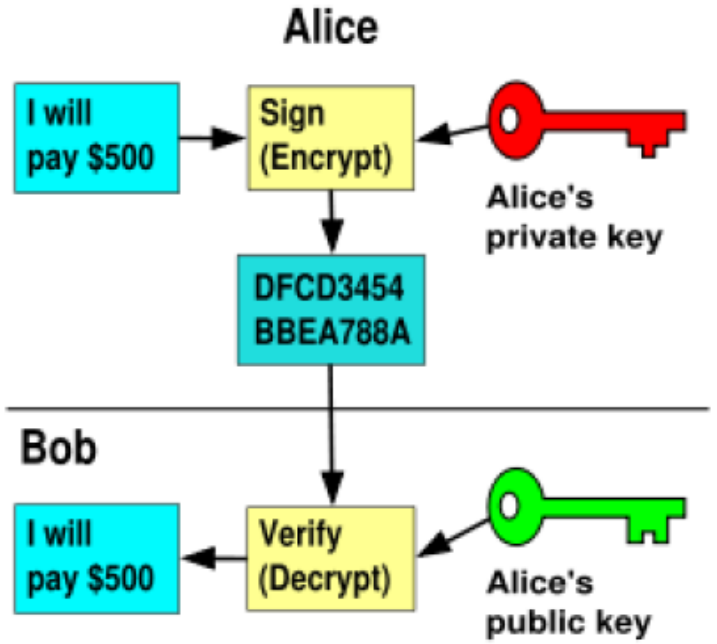
- Skupina kryptografických metod, ve kterých se pro šifrování a dešifrování používají *odlišné* klíče
  - jedna část se používá pro šifrování zpráv (a příjemce zprávy ani tuto část nemusí znát)
  - druhá pro dešifrování (a odesilatel šifrovaných zpráv ji zpravidla nezná)
- Nejběžnější verzí asymetrické kryptografie je využívání tzv. veřejného a soukromého klíče
  - šifrovací klíč je veřejný, majitel klíče ho volně uveřejní, a kdokoli jím může šifrovat jemu určené zprávy
  - dešifrovací klíč je soukromý, majitel jej drží v tajnosti a pomocí něj může tyto zprávy dešifrovat
- Kromě možnosti pro utajení komunikace se asymetrická kryptografie používá také pro elektronický podpis

laboratory Gerstner

Example I



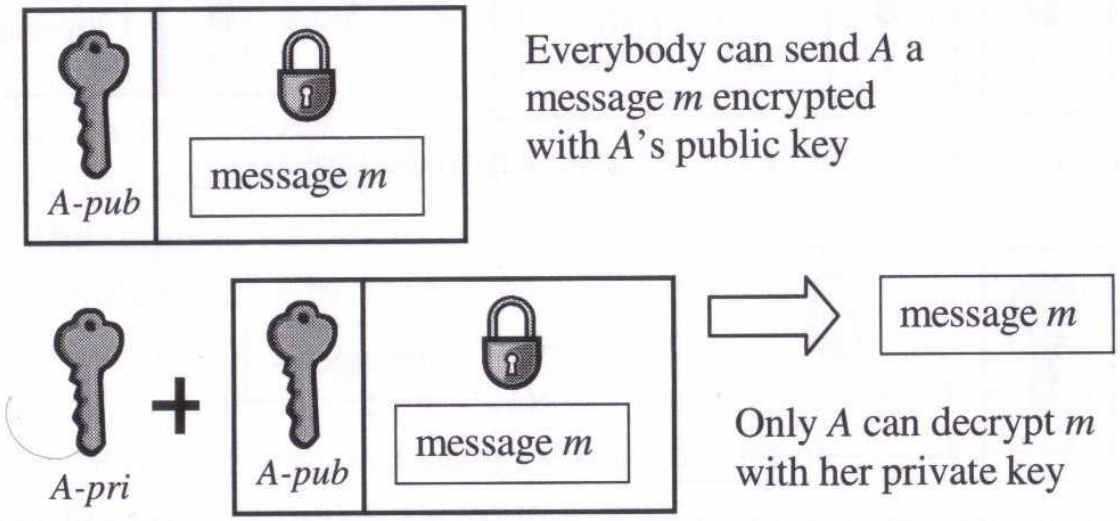
A big random number is used to make a public-key pair



Using a private key to encrypt (thus signing) a message; anyone can check the signature using the public key.
Validity depends on private key security.

Example II

$D(E(m,k_{A\text{-}pub}),k_{A\text{-}pri}) = m$ and conversely,
$D(E(m,k_{A\text{-}pri}),k_{A\text{-}pub}) = m.$



Elektronický podpis

- **Elektronický podpis** (někdy také digitální podpis) je certifikát, který plně nahrazuje vlastnoruční podpis při elektronické komunikaci.
- Umožňuje ověřit **integritu** podepsaného dokumentu (tj. zda v dokumentu něco nechybí nebo nepřebývá).
- Funguje na principu **asymetrického šifrování.**
- **Z podpisovaného dokumentu je nejprve vytvořen otisk (tzv. hash).**
- Tento otisk je poté zašifrován tajným klíčem autora podpisu.
- Šifrování je poměrně rychlé a hardwarově nenáročné.

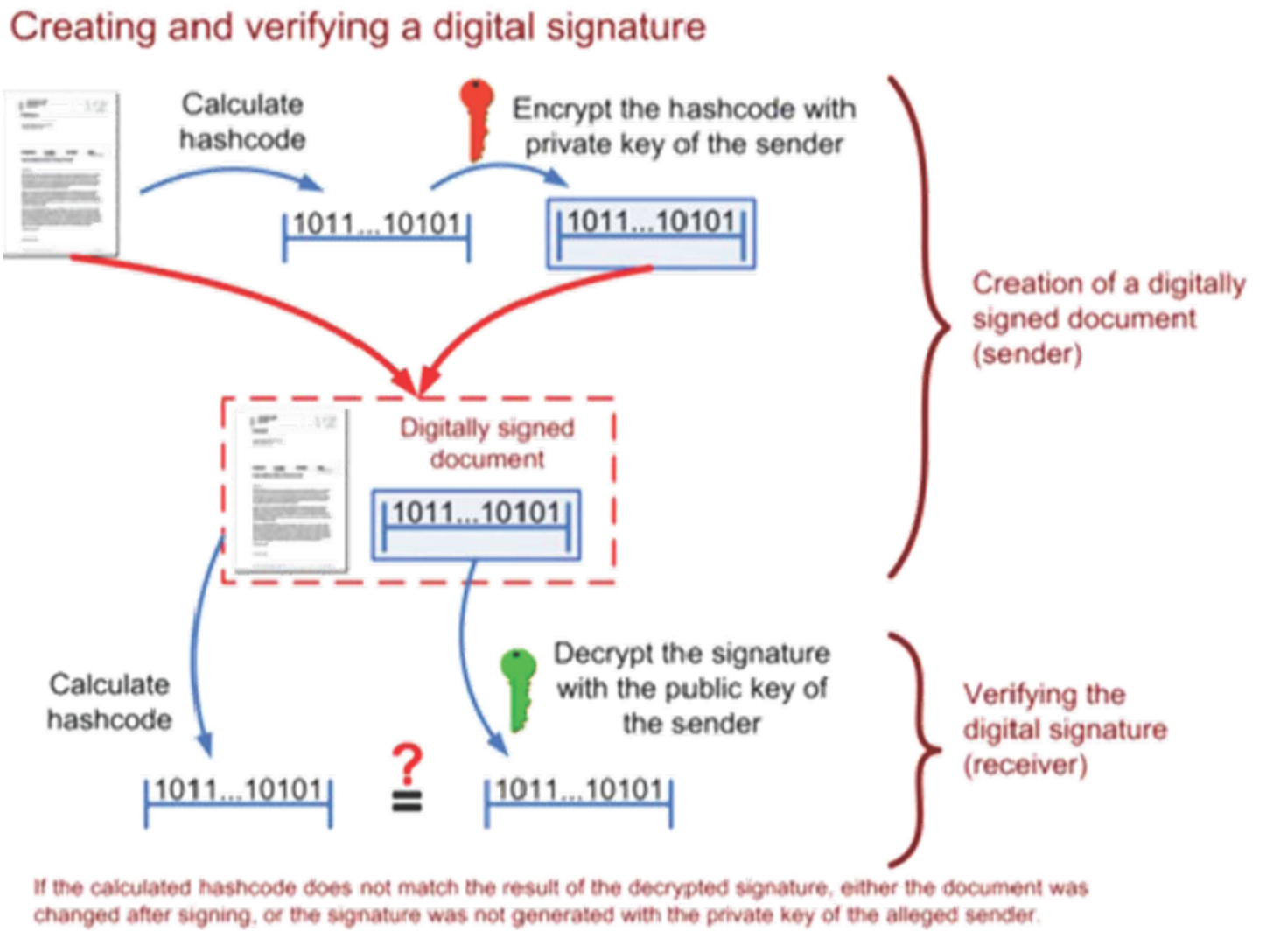
laboratory Gerstner

Princip funkce elektronického podpisu

- Před posláním
  - Ze zprávy je spočítán pomocí hašovací funkce hash.
  - Hash je zašifrován privátním klíčem.
  - Zpráva je poslána se zašifrovaným hashem.

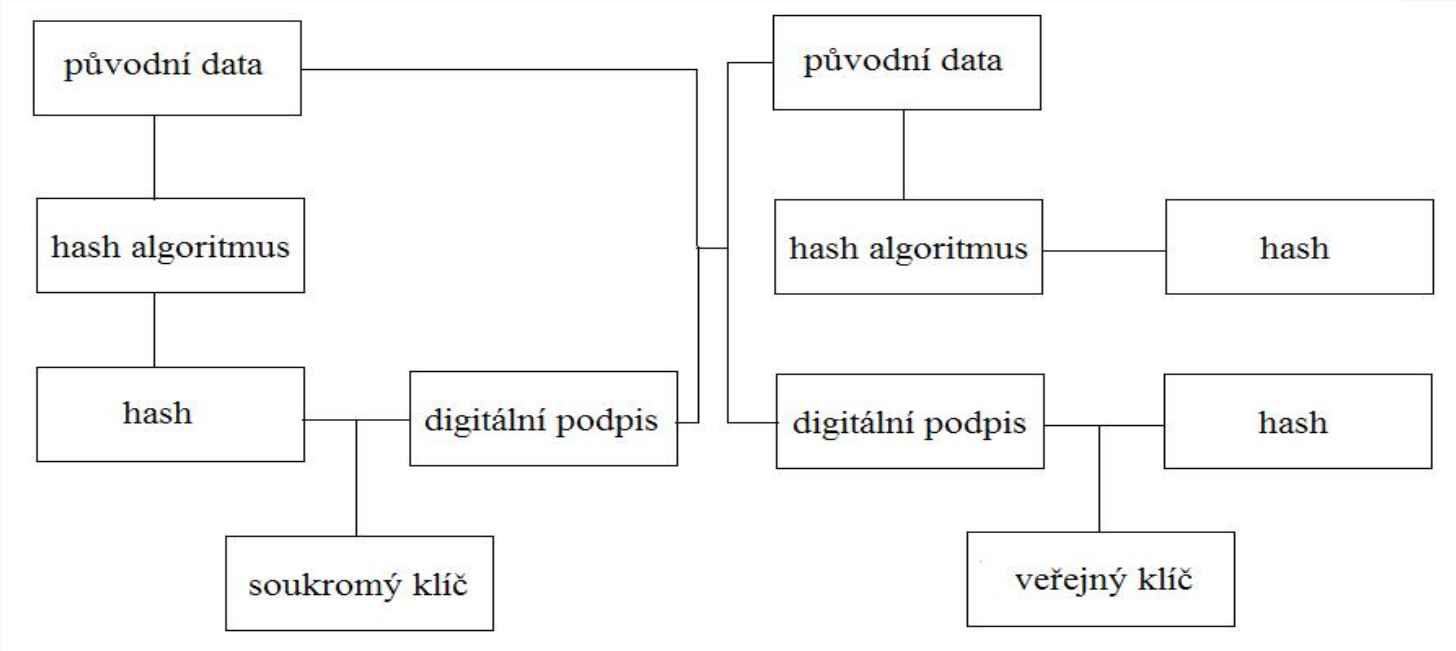
- Po poslání
  - Zašifrovaný hash je dekódován pomocí veřejného klíče.
  - Ze zprávy je vypočítán hash a porovnán s dekódovaným.

Digital signature

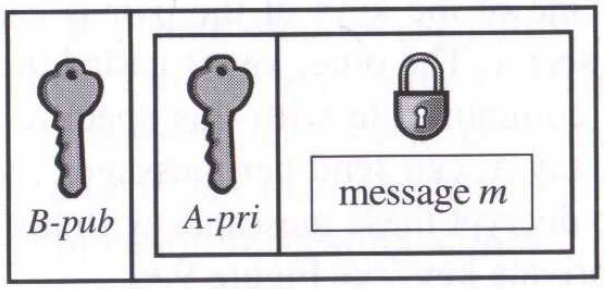


Elektronický podpis

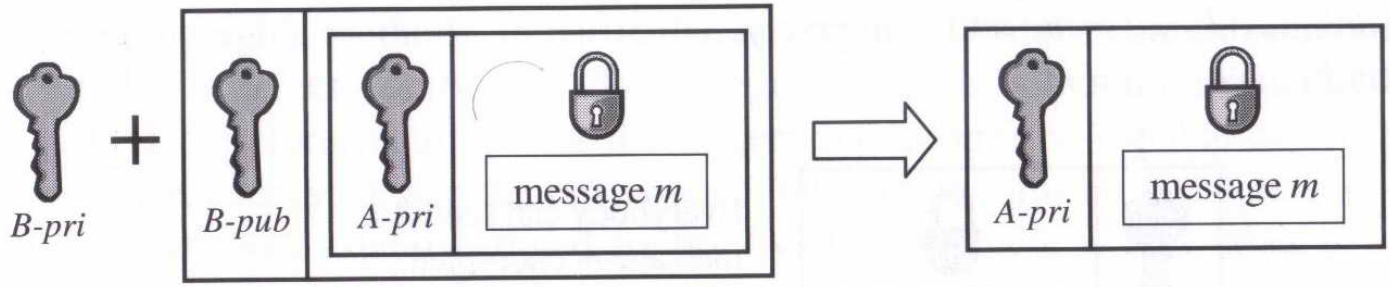
- Algoritmy pro vytvoření digitálního podpisu
  - Asymetrické kryptovací algoritmy s veřejným klíčem, nejčasteji RSA (Rivest-Shamir-Adleman) a DSA (Digital Signature Algorithm)
  - Bezpečné kryptografické jednocestné algoritmy (hashovací funkce), nejčastěji MD5 (Message Digest 5) spolu s RSA a SHA (Secure Hash Algorithm) spolu s DSA



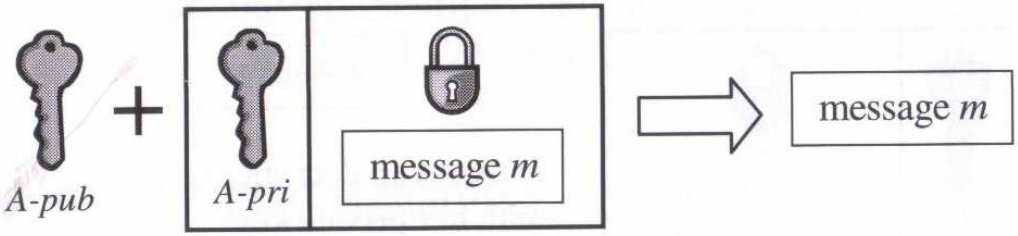
Digital Signature



A sends *B* a message *m* which is first encrypted with *A*'s private key and then with *B*'s public key



B first uses his private key and then *A*'s public key to decrypt *m*



Hašovací funkce

- Předpis pro výpočet kontrolního součtu (haše) ze zprávy.
- Slouží ke kontrole integrity dat, k rychlému porovnání dvojice zpráv, indexování, vyhledávání apod.
- Formálně je to funkce h, která převádí vstupní posloupnost bitů na posloupnost pevné délky n bitů.

$$h : D \; ! \; R, \; |D| > |R|$$

- Kolize
  - Dvojice vstupních dat (x, y), které mají stejný hash.

$$h(x) = h(y)$$

Požadavky na hashovací funkci

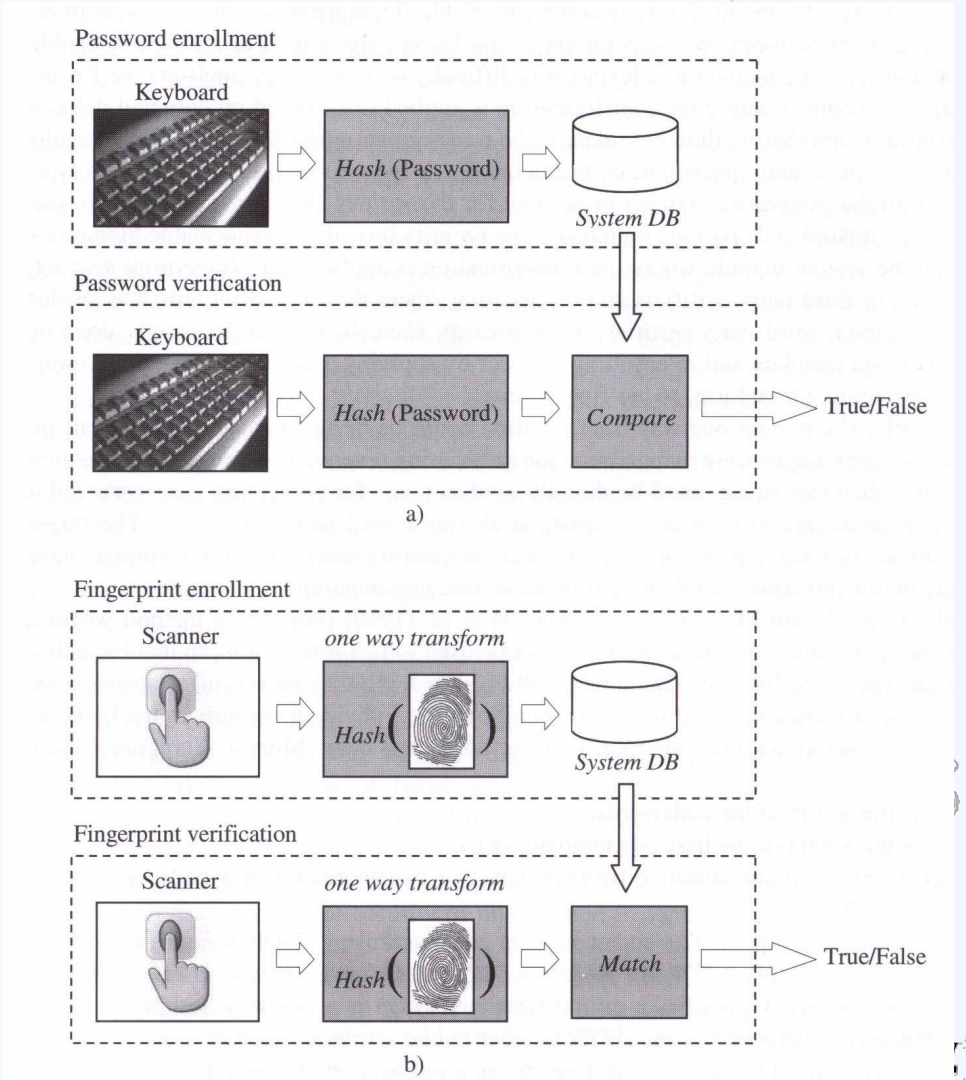
- **Hashovací funkce** je předpis pro výpočet kontrolního součtu (hashe) ze zprávy či většího množství dat
- Nejdůležitější je následující trojice vlastností. Obtížností se v tomto kontextu myslí výpočetní složitost
  - Odolnost vůči získání předlohy. Pro daný hash c je obtížné spočítat x takové, že $h(x)=c$. (Hashovací funkce je jednosměrná.)
  - Odolnost vůči získání jiné předlohy. Pro daný vstup x je obtížné spočítat y takové, že $h(x)=h(y)$
  - Odolnost vůči nalezení kolize. Je obtížné systematicky najít dvojici vstupů (x,y), pro které $h(x)=h(y)$
  - Nekorelovatelnost vstupních a výstupních bitů, kvůli znemožnění statistické kryptoanalýzy
  - Odolnost vůči skoro-kolizím. Je obtížné nalézt x a y taková, že $h(x)$ a $h(y)$ se liší jen v malém počtu bitů
  - Lokální odolnost vůči získání předlohy. Je obtížné najít i jen část vstupu x ze znalosti $h(x)$.

Používané Hašovací funkce

- MD-5
  - Od srpna 2004 je veřejně znám postup nalezení kolizí a vstupní data se dokonce jen málo odlišují.
- SHA-1
  - V únoru 2005 byl zveřejněn objev algoritmu, který umožňuje nalézt kolizi podstatně rychleji než hrubou silou.
  - Výpočetní náročnost je ale stále mimo současnou techniku.
- SHA-2
  - Dosud považována za spolehlivou.

Cancelable biometrics

- Can not be used in another application
- If template or DB comprimised, a new record is issued
- Altering DB record impossible, a template is digitally signed, or encryption key is stored in a template
- Main idea
  - Non-invertible transform
  - One way hash fce
- SOFT matching, passwds are the same.



Example of securing coercion

- Users are forced to identify at gunpoint
- ATM scenario
- Default finger (right middle), Panic finger (right index)
- Willful withdrawal -> default finger
- Coerced withdrawal -> panic finger
  - ATM dispense currency with special invisible ink
  - Acquire snapshot of surveillance video
  - Inform authorities

laboratory Gerstner