evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

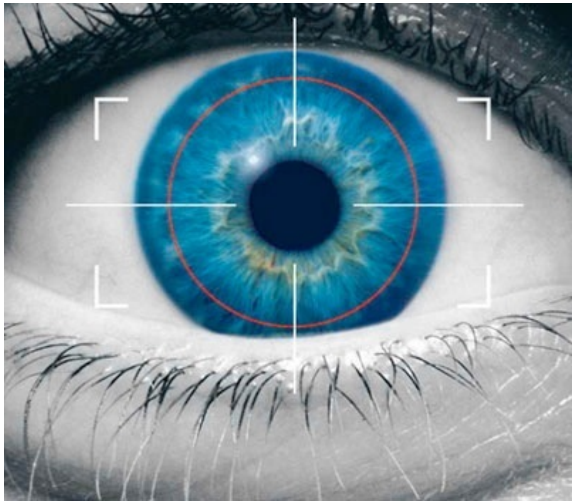
Biometric Cryptography - Mobile Application Viewpoint

Tomáš Rosa
Raiffeisenbank, a.s.
crypto.hyperlink.cz

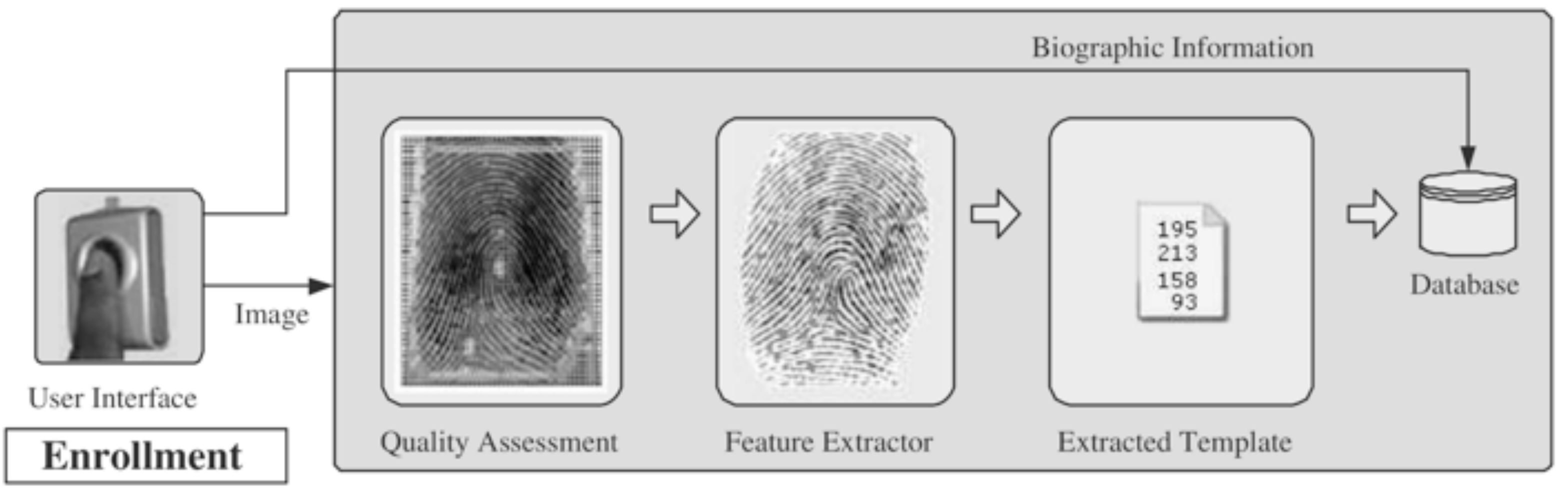
biocryptography, Brno, 2014

Biometric Identification/Verification

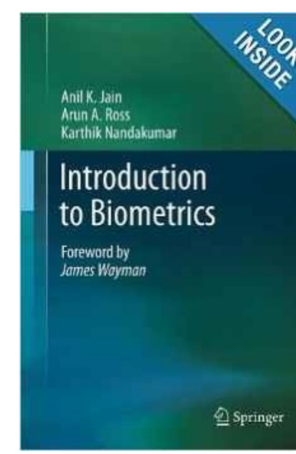
...automated establishment of the human identity based on their physical or behavioral characteristics.



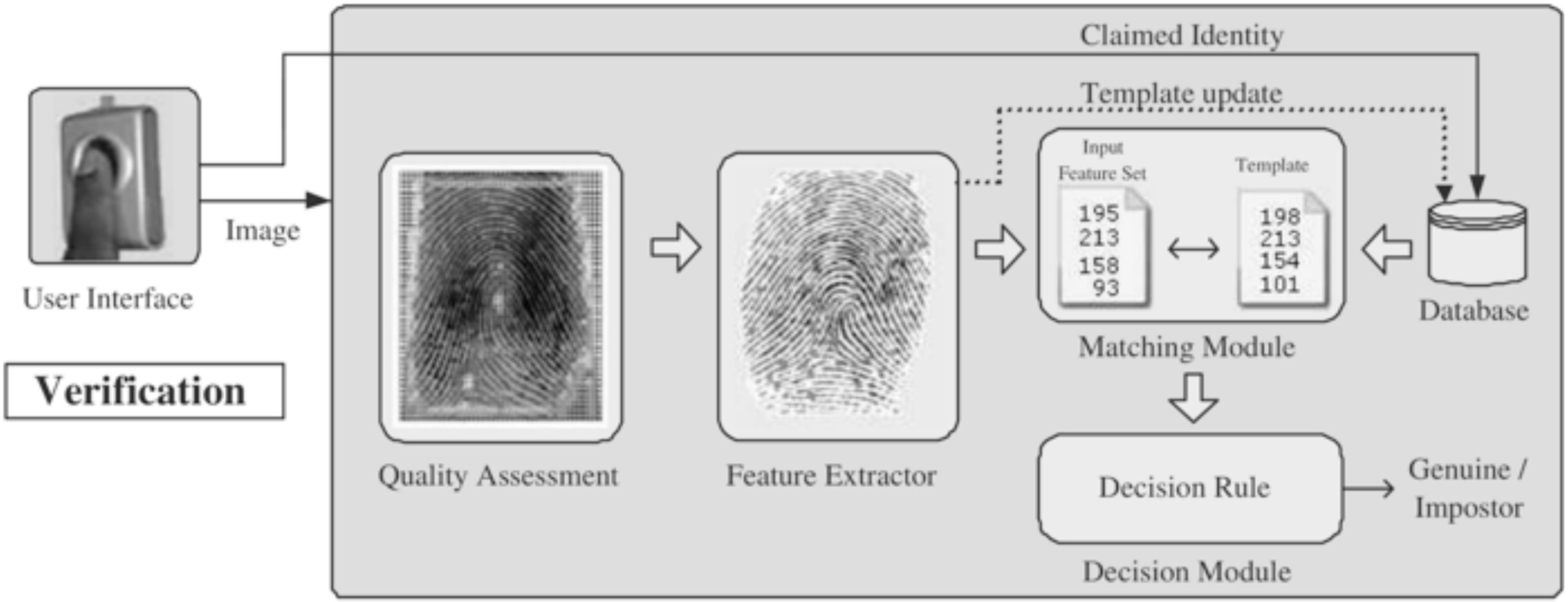
Enrolment Phase



Jain, Ross, Nandakumar: *Introduction to Biometrics*, Springer, 2011

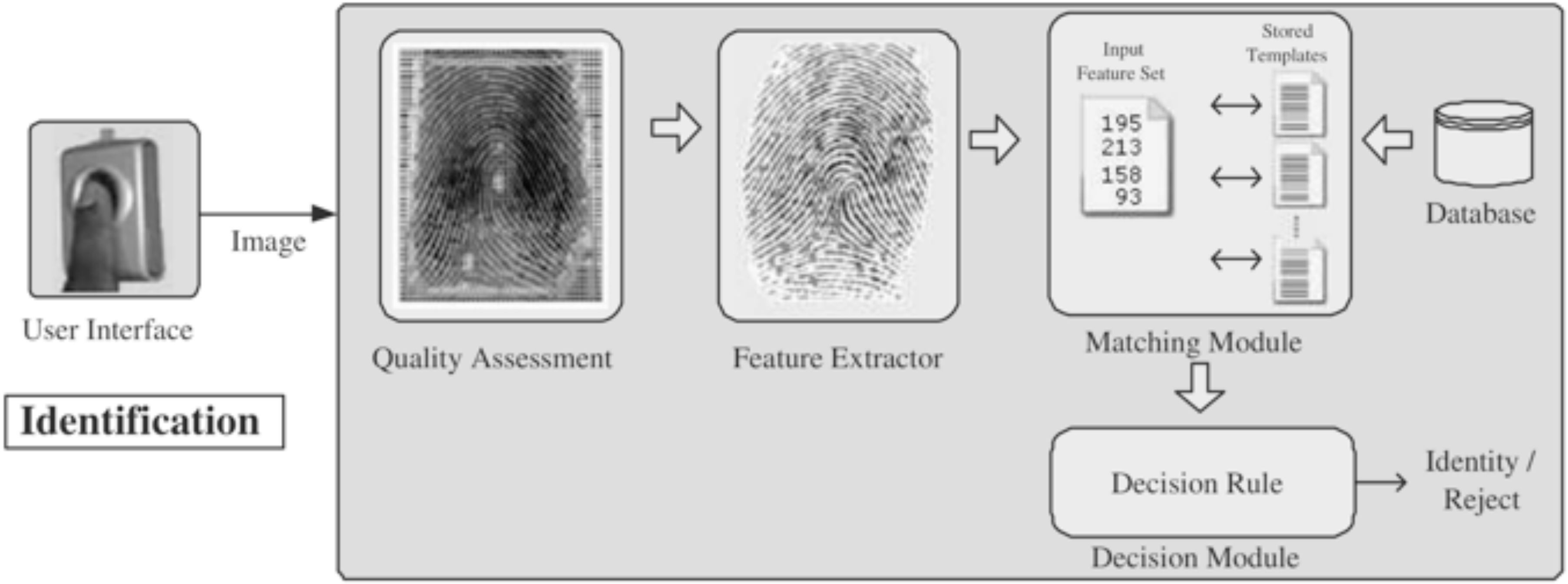


Verification (1 : 1)



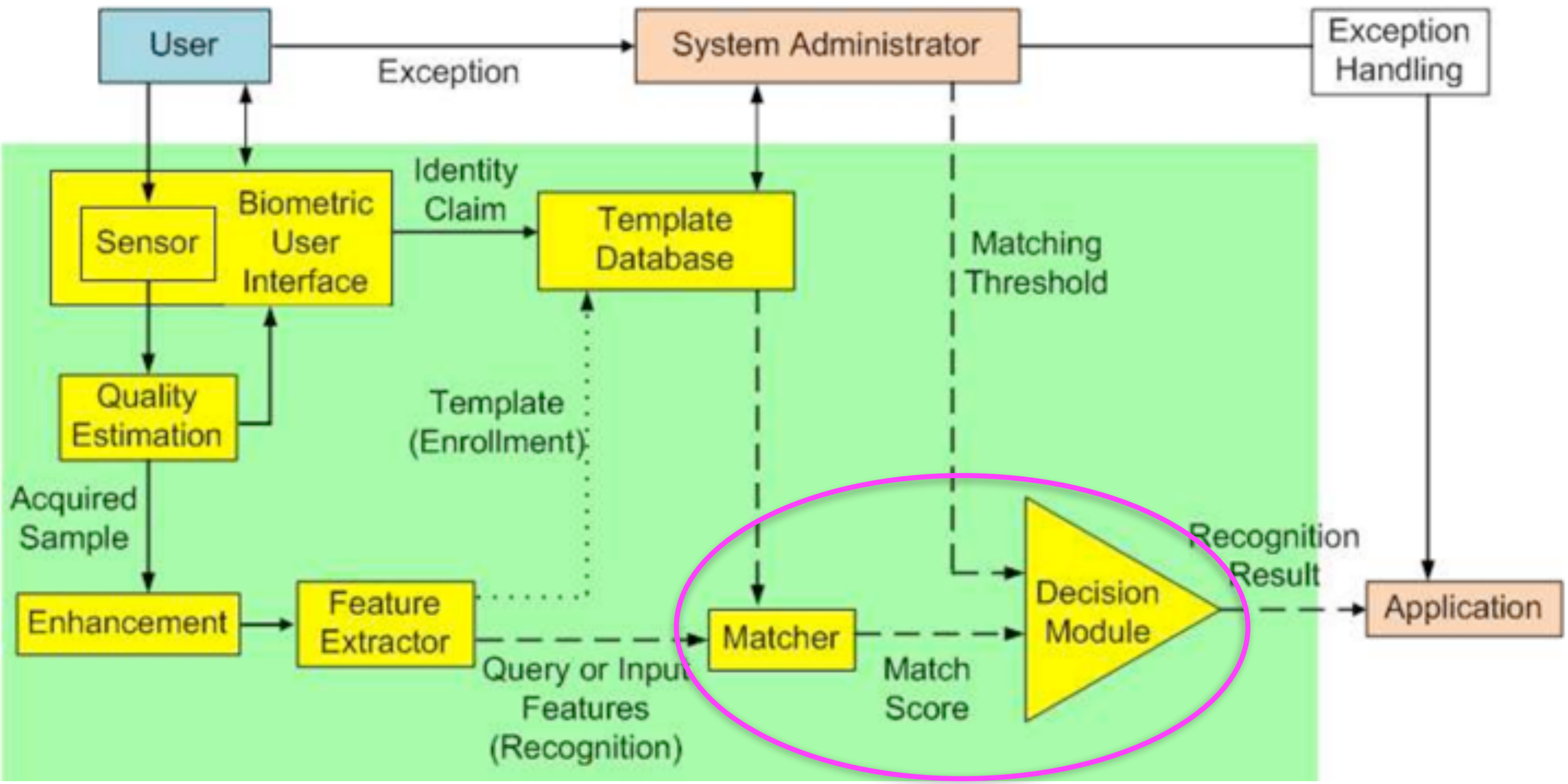
Jain, Ross, Nandakumar: *Introduction to Biometrics*, Springer, 2011

Identification (1 : N)



Jain, Ross, Nandakumar: *Introduction to Biometrics*, Springer, 2011

Biometric System Topology

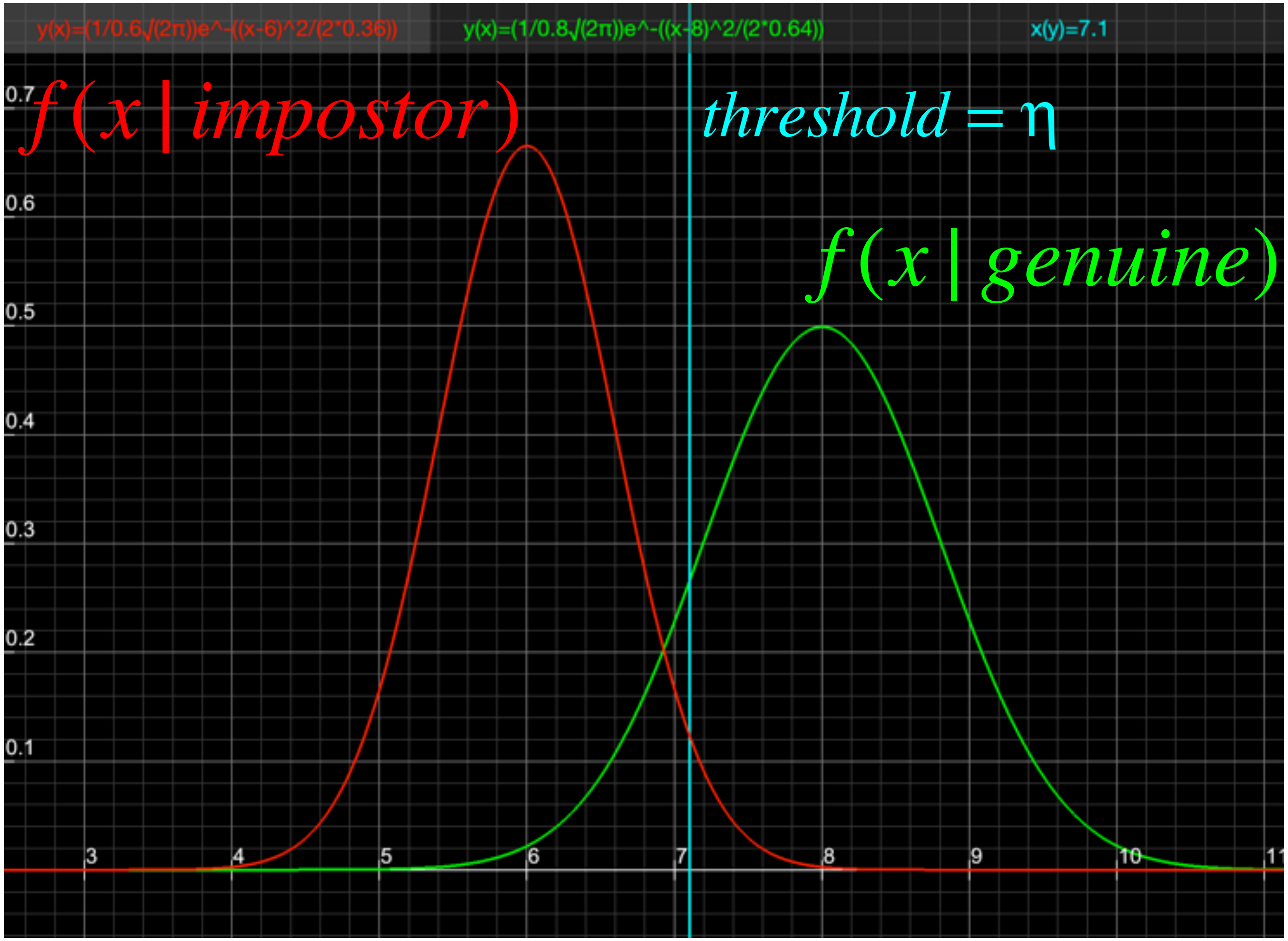


Jain, Ross, Nandakumar: *Introduction to Biometrics*, Springer, 2011

Match Score

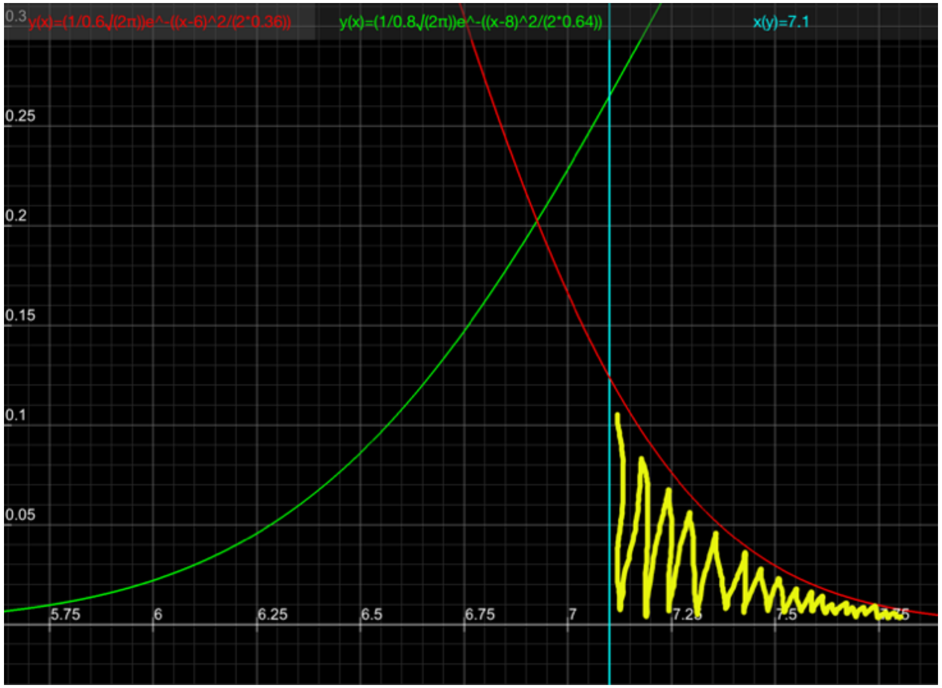
- It would be nice if we had simple true/ false result.
  - As in conventional crypto.
  - But we cannot...
- All we have is a random variable X that follows two conditional distributions.
  - $f(x \mid \text{impostor})$
  - $f(x \mid \text{genuine})$

Match Score Evaluation



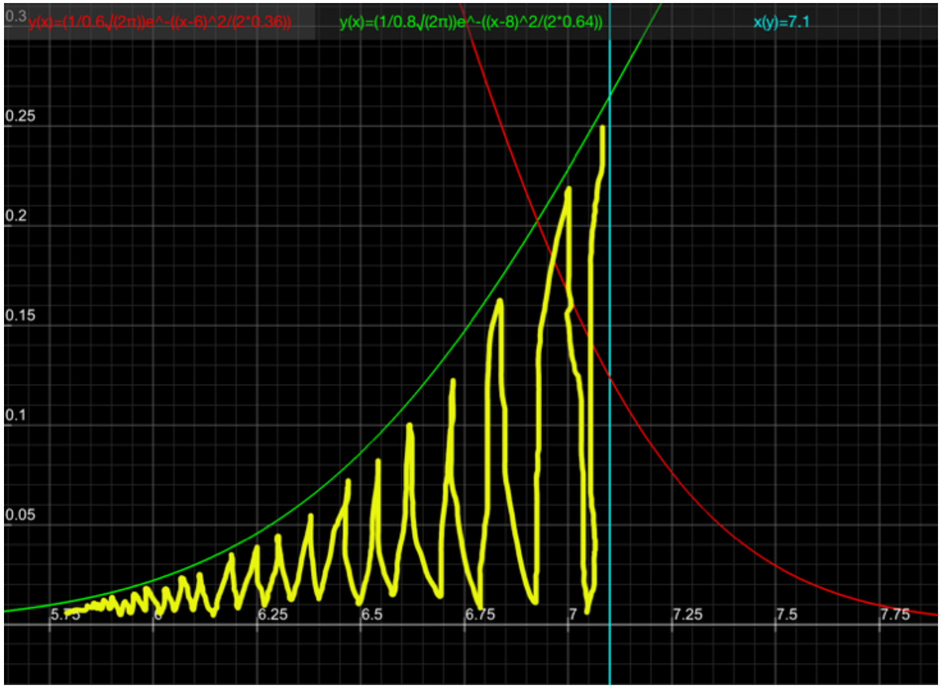
False Acceptance Rate

$$FAR = \int_{\eta}^{\infty} f(x \mid impostor)\,dx$$

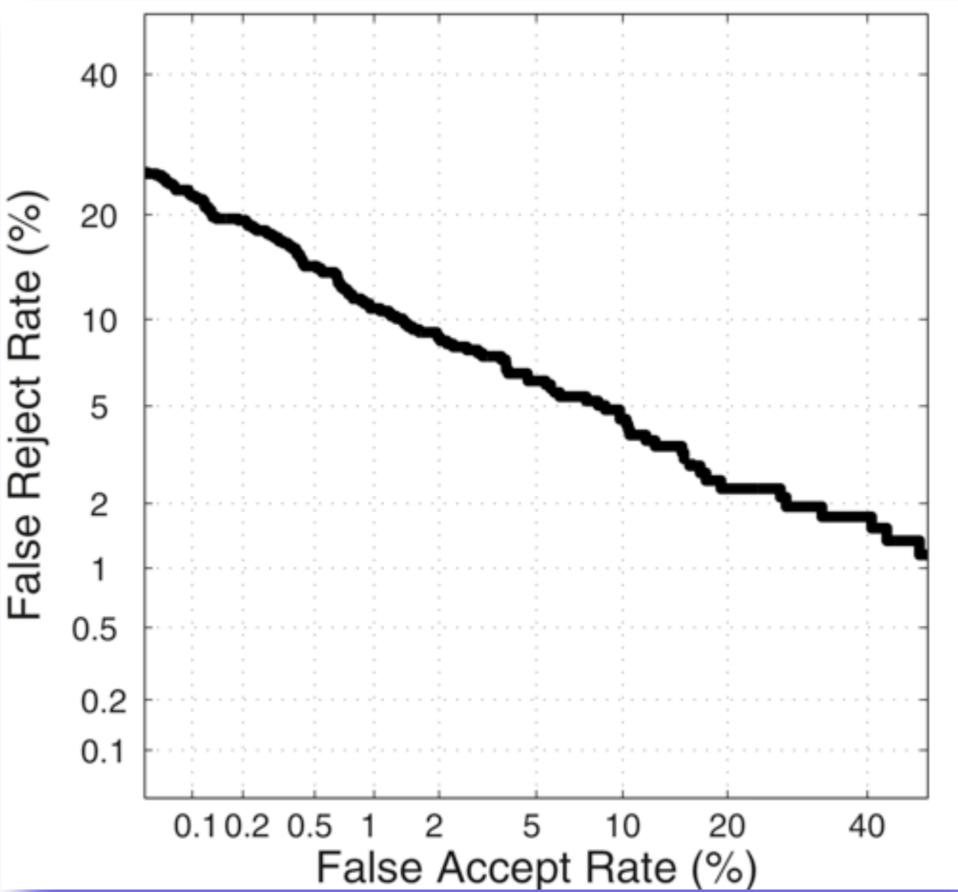


False Rejection Rate

$$FRR = \int_{-\infty}^{\eta} f(x \mid genuine)\,dx$$



Real DET Curve



Detection
Error
Tradeoff

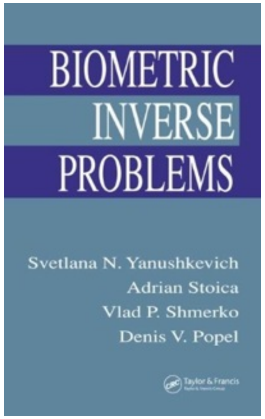
Jain, Ross, Nandakumar, Springer 2011

Contrasting Design Approach

- Classic cryptography
  - infeasible mathematical problems
- Quantum cryptography
  - intractable physical problems
- Biometric identification
  - statistical signal analysis and pattern recognition
  - intractability is usually ***not*** the prime concern
  - we hope the Mother Nature complexity *somehow* guarantees the security

BIO Brute Force Attack

- Randomly generate plausible circa 1/FAR samples and put them to the test.
  - Also termed “Zero-Effort”, denoting that the attacker makes no special effort to imitate the original person characteristic.
- Synthetic samples generation is quite feasible today.



Cryptanalysis-Like Attacks

- Usually a variant of "Hill-Climbing" denoting the attacker iteratively improves the BIO sample data based on:
  - scoring feedback *(side channels)*
  - stolen template *(pre-image attacks)*
  - independent template trained from intercepted BIO samples *(correlation attacks)*
  - known scoring anomaly *(differential analysis. etc.)*
  - implementation faults *(general hacking)*

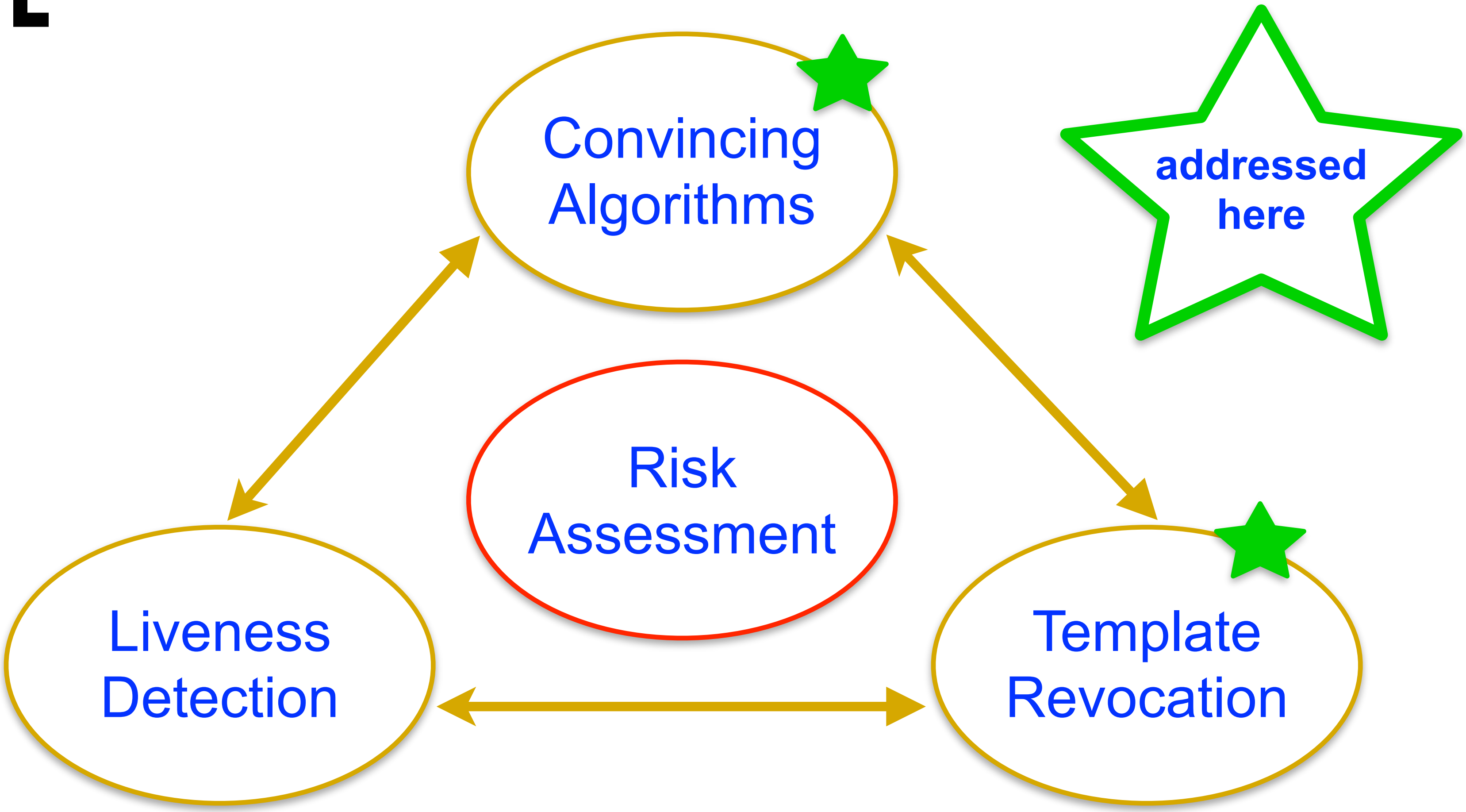
Spoofing

- *The process of defeating a biometric system through the introduction of fake biometric samples.*
  - *(Schuckers, Adler et al., 2010)*
- Particular modus operandi on how to deploy the attacking data vectors.
  - Can be seen as being orthogonal to the aforementioned hill-climbing attacks.

Voice Biometrics Spoofing

- Spoofing techniques are, however, not "just helpers" as they are interesting on their own:
  - Text-To-Speech Synthesis
  - Voice Conversion
  - Artificial Signals

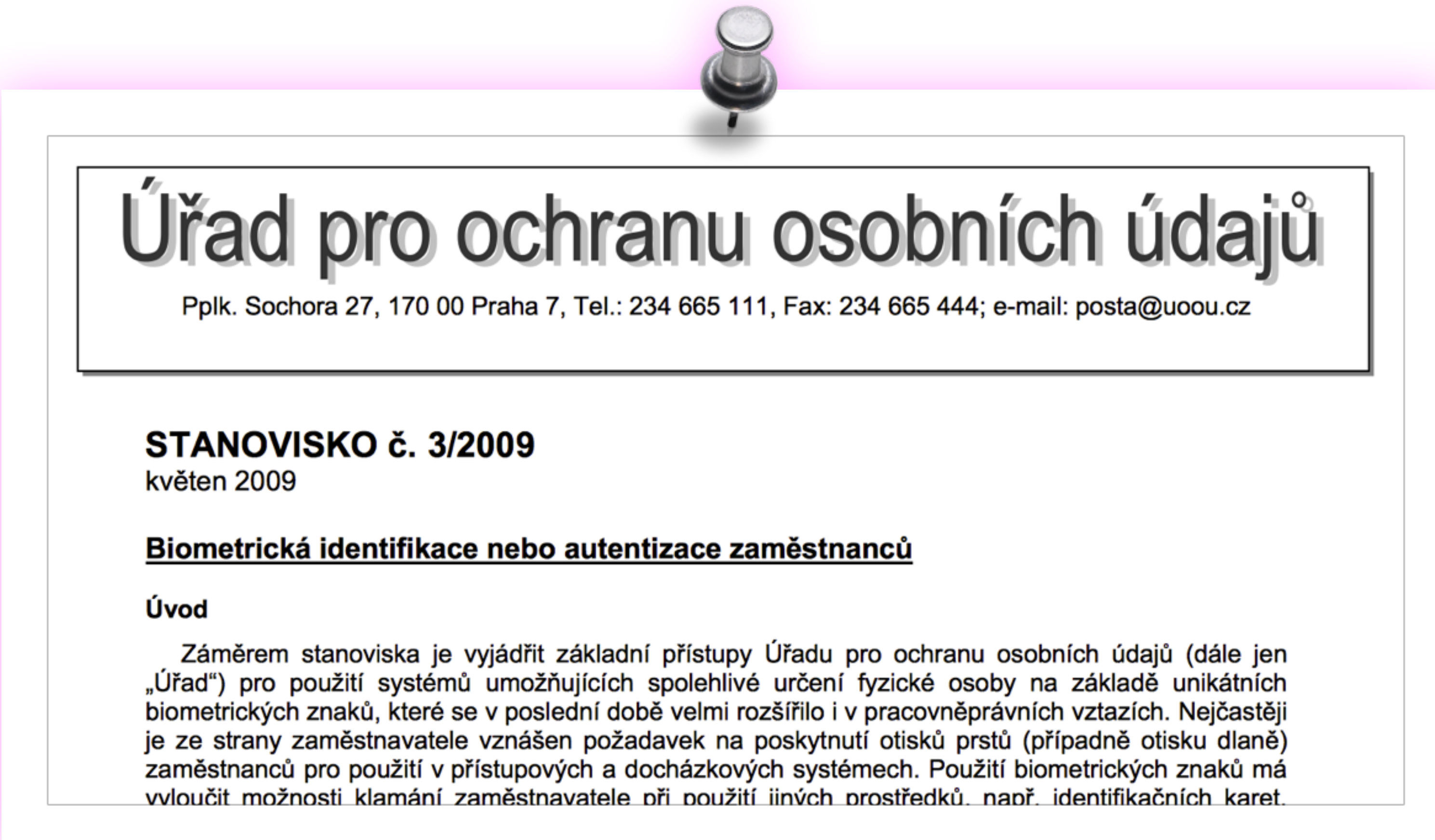
Open Problems



Biometrics In Mobile App

- Let’s say we want to enhance a mobile banking application by biometrics.
- ...three-factor authentication by:
    I) something to have (device key)
    II) something to know (PIN)
    III) something to be (BIO sample)

Reflecting Privacy Protection



Úřad pro ochranu osobních údajů

Pplk. Sochora 27, 170 00 Praha 7, Tel.: 234 665 111, Fax: 234 665 444; e-mail: posta@uoou.cz

STANOVISKO č. 3/2009
květen 2009

Biometrická identifikace nebo autentizace zaměstnanců

Úvod

Záměrem stanoviska je vyjádřit základní přístupy Úřadu pro ochranu osobních údajů (dále jen „Úřad") pro použití systémů umožňujících spolehlivé určení fyzické osoby na základě unikátních biometrických znaků, které se v poslední době velmi rozšířilo i v pracovněprávních vztazích. Nejčastěji je ze strany zaměstnavatele vznášen požadavek na poskytnutí otisků prstů (případně otisku dlaně) zaměstnanců pro použití v přístupových a docházkových systémech. Použití biometrických znaků má vyloučit možnosti klamání zaměstnavatele při použití jiných prostředků, např. identifikačních karet.

Privacy Protection Conclusion

- There is a strong preference of biometric systems such that:
  - they do not process biometric samples left unintentionally
  - they do not store biometric template in one central database

addressed here

Local Templates

- We want to process the biometric data strictly locally in the mobile device.
  - So the bank does not store the precious BIO templates of its clients.
- Furthermore, we want to leverage the existing mechanism of distributed implicit PIN verification via (H)OTP.
  - cf. "*The Decline and Dawn of Two-Factor Authentication on Smart Phones*", ISS 2012

Naive Approach

```
sample = get_biometric_data();

if (match(sample, template) > eta)
    continue_with_authentication();
else
    abort_authentication();
```

Recall ATA

Definition. *Let the After-Theft Attack (ATA) be any attacking scenario that assumes the attacker has unlimited physical access to the user's smart device.*

- Imagine somebody steals your mobile phone…
- Despite being a really obvious threat, it is way too often neglected in contemporary applications.

- By a robbery, the attacker can even get access to unlocked screen or a paired computer, hence receiving another considerable favour!

Naive Approach vs. ATA

```
sample = get_biometric_data();

if (match(sample, template) > eta)
   continue_with_authentication();
else
   abort_authentication();
```

Naive Approach vs. ATA

```
sample = get_biometric_data();

if (match(sample, template) > eta)
    continue_with_authentication();
else
    abort_authentication();
```

bypassed!

Naive Approach vs. ATA

```
sample = get_biometric_data();

if (match(sample, template) > eta)
    continue_with_authentication();
else
    abort_authentication();
```

bypassed!

stolen!

Intermezzo

Recall how we process the PIN in mobile apps:

i) unlock a *PIN_key* by the PIN

ii) let *MK* = *KDF*(*PIN_key*, *device_key*)

iii) verify *MK* with the bank using conventional crypto protocols

...distributed implicit PIN verification.

Intermezzo

PIN_key is shared with the bank (not the PIN!)

Recall how we process the PIN in mobile apps:

i) unlock a *PIN_key* by the PIN

ii) let *MK* = *KDF*(*PIN_key*, *device_key*)

iii) verify *MK* with the bank using conventional crypto protocols

...distributed implicit PIN verification.

Adding the BIO Factor

Is there something like "*BIO_key*"?

We would have:

i) unlock the *PIN_key* by the PIN
ii) unlock the *BIO_key* by the user's BIO
iii) let *MK* = *KDF*(*PIN_key*, *BIO_key*, *device_key*)
iv) verify *MK* with the bank using conventional crypto protocols

Adding the BIO Factor

Is there something like "*BIO_key*"?

We would have:

i) unlock the *PIN_key* by the PIN
ii) unlock the *BIO_key* by the user's BIO
iii) let *MK* = *KDF*(*PIN_key*, *BIO_key*, *device_key*)
iv) verify *MK* with the bank using conventional crypto protocols

Again, BIO_key is shared with the bank, not a BIO template

Cryptography Exactness

$$\text{Let } y = AES_K(x) \text{ for a random } K.$$

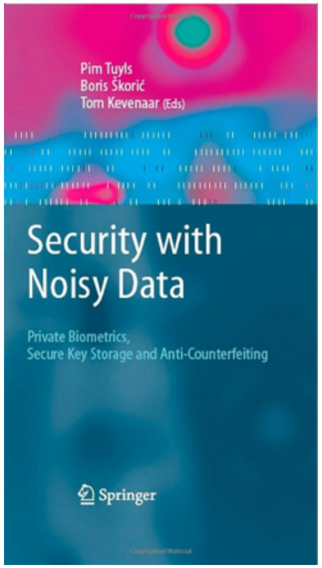
$$\text{Then } AES_K^{-1}(y) = x, \text{ while}$$

$$AES_{K \oplus 1}^{-1}(y) \neq x \text{ (probability} \approx 1).$$

- The better the algorithm is the more randomized response we get for even one-bit error.

Biometrics Fuzziness

- We seldom get the same data in the subsequent scans of the very same person.
  - Actually, this is usually a clear sign of a spoofed sample.
- To overcome this (intra-user) variability, we can employ the *biometric cryptography*.



BIO Cryptography

- Well, in 90's, there was a lot of alchemy in there.
  - Same as in crypto before C. E. Shannon in 1948 - 1949.
- Nowadays, it works hard towards a respected science.
- ...or how to deal with noisy data in cryptographic transformations.
  - These ideas go beyond the scope of biometrics. Quantum crypto or PUFs are further examples…
  - We can see the biometric cryptography as combining both feature quantization and classification into one "convolved" protocol.

Our Illustrative Approach

- We employ BIO cryptography to cope with ATA threat in the mobile app.
  - On behalf of this, we discuss the key concepts of these algorithms and protocols.

Error-Correcting Code C

Let (F,ρ) be a metric space, $\rho: F \times F \to [0,\infty)$.

translation invariant metric: $\rho(x,y) = \rho(0,x-y)$

Error correcting code is $C \subset F, C = \{c_1, c_2, \ldots\}$.

$decode: F \to C$

t-error correction capability:

Let $\rho(c_i, y) \le t$, then $decode(c_i) = decode(y) = c_i$.

We assume $decode()$ always returns
a (possibly wrong) codeword.

Metric For the Biometrics

- Let the extracted biometric features be expressible as an element of (***F***, *ρ*).
  - Let also the *ρ*-distance measures the (dis)similarity of the two BIO samples.
    - We follow the *Fuzzy Commitment* by Juels and Wattenberg scheme that is a very good teaching example, since 1999.
    - It was (i.a.) generalised by Dodis et al. (2004) as *Fuzzy Extractor* based on *Secure Sketch*.
    - A well structured experiment exposing a particular ECC design to work with the iris code is by Hao et al. (2005).

ECC Theory DO's and DON'Ts

- Recall, for ECC, we have solid proofs of guaranteed *random error correction capabilities*.
  - However, this is not the same as proofs of guaranteed ***correlated*** *error correction* ***in****capabilities*.
- We need to combine low-level equation inspection together with overall statistics to get the assurance we want.

Enrolment

i) randomly choose $c_{key} \in \mathbf{C} \subset \mathbf{F}$

ii) get BIO features vector $w \in \mathbf{F}$

iii) let $\xi = w - c_{key}$

iv) let $BIO\_key = hash(c_{key})$

v) template = (ξ)

Enrolment

i) randomly choose $c_{key} \in \mathbf{C} \subset \boldsymbol{F}$

ii) get BIO features vector $w \in \boldsymbol{F}$

iii) let $\xi = w - c_{key}$

iv) let $BIO\_key = hash(c_{key})$

v) template = (ξ)

More involved entropy extractors can be used here…

Verification

i) get BIO features vector $w' \in \boldsymbol{F}$

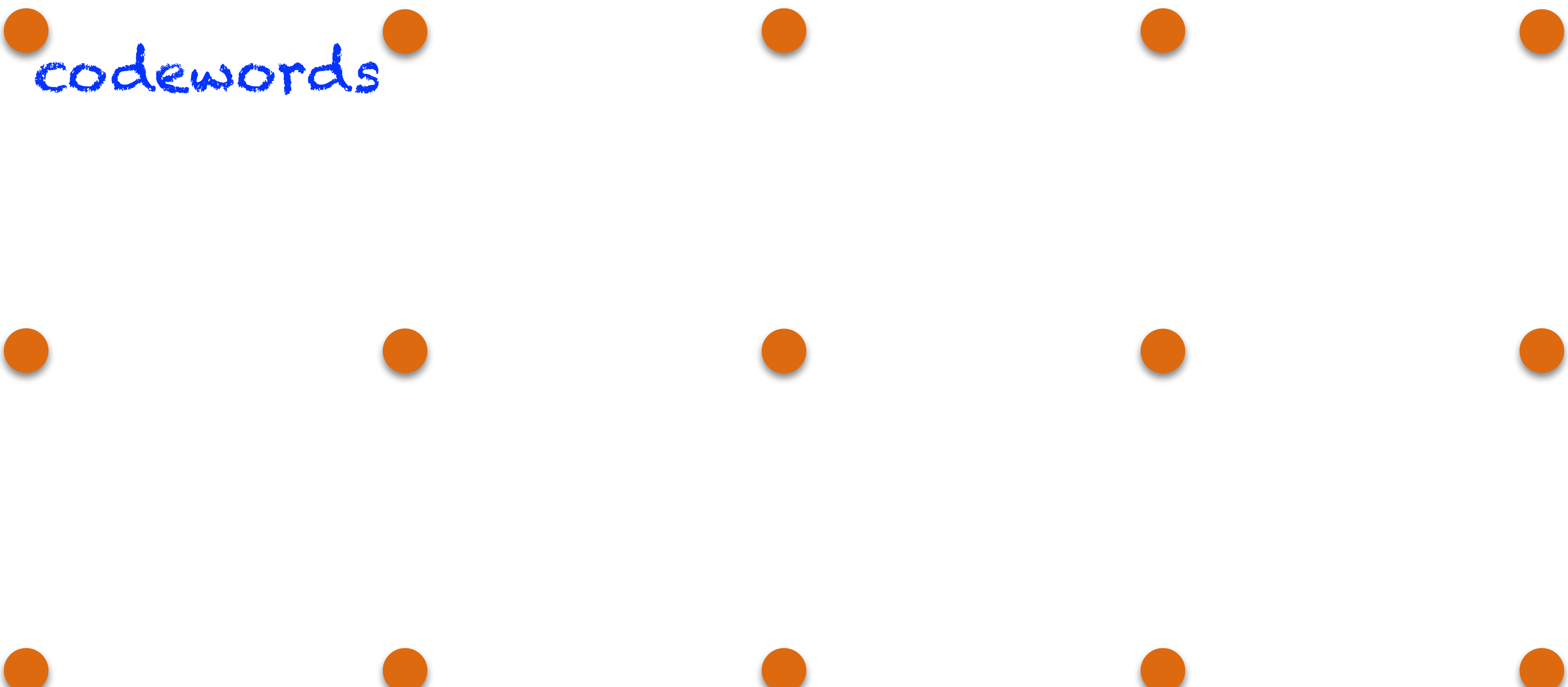
ii) let $y = w' - \xi$

iii) let $c_{key}' = decode(y)$

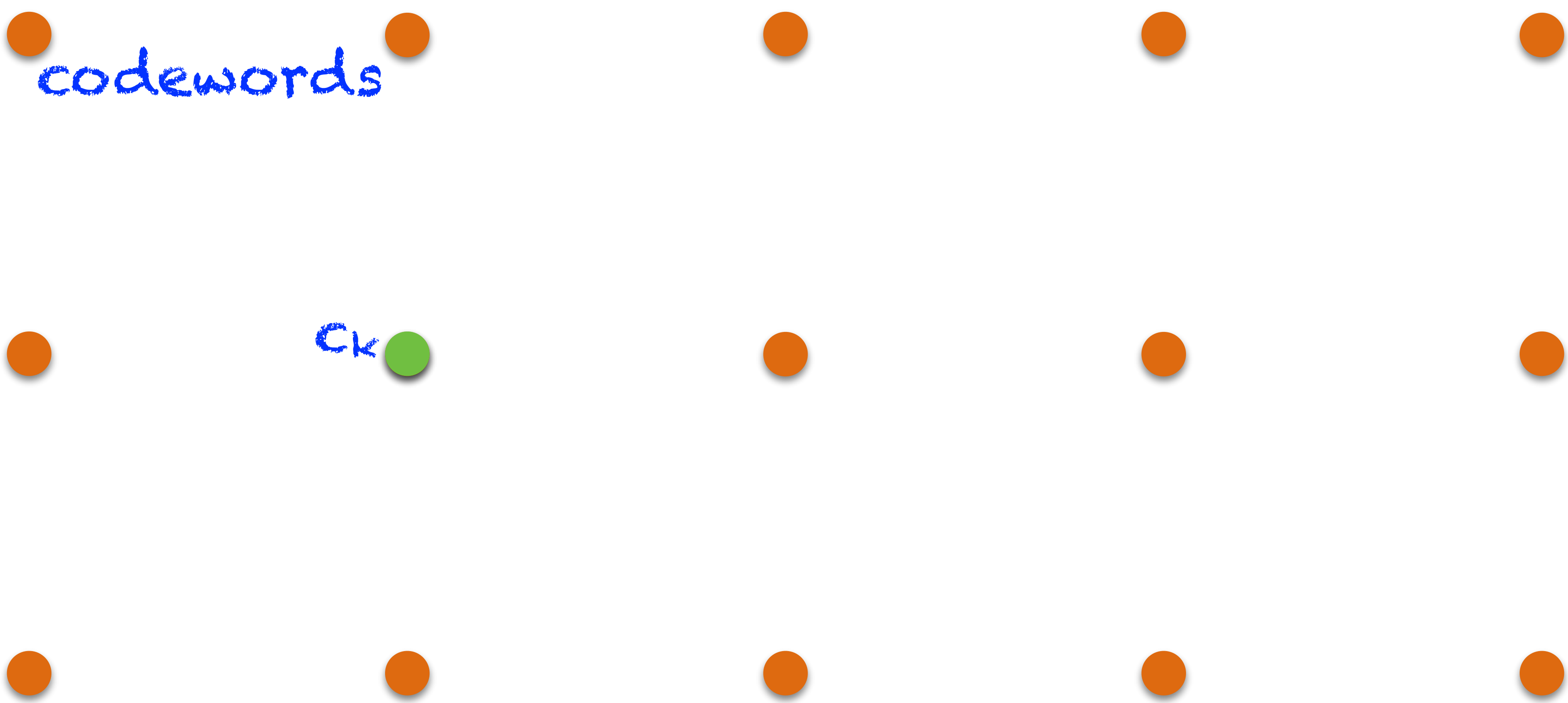
iv) let $BIO\_key' = hash(c_{key}')$

v) try to use BIO_key' in the protocol above

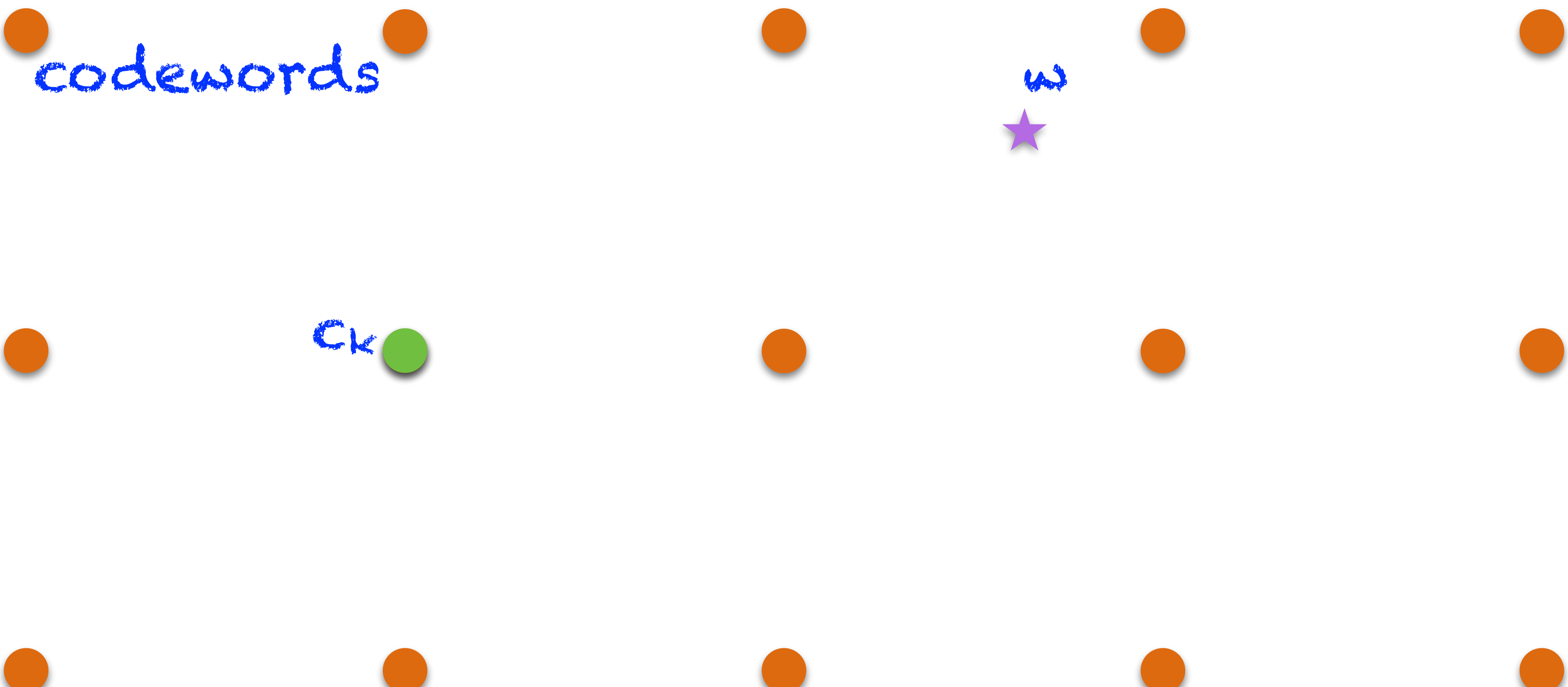
Core Principle Illustrated



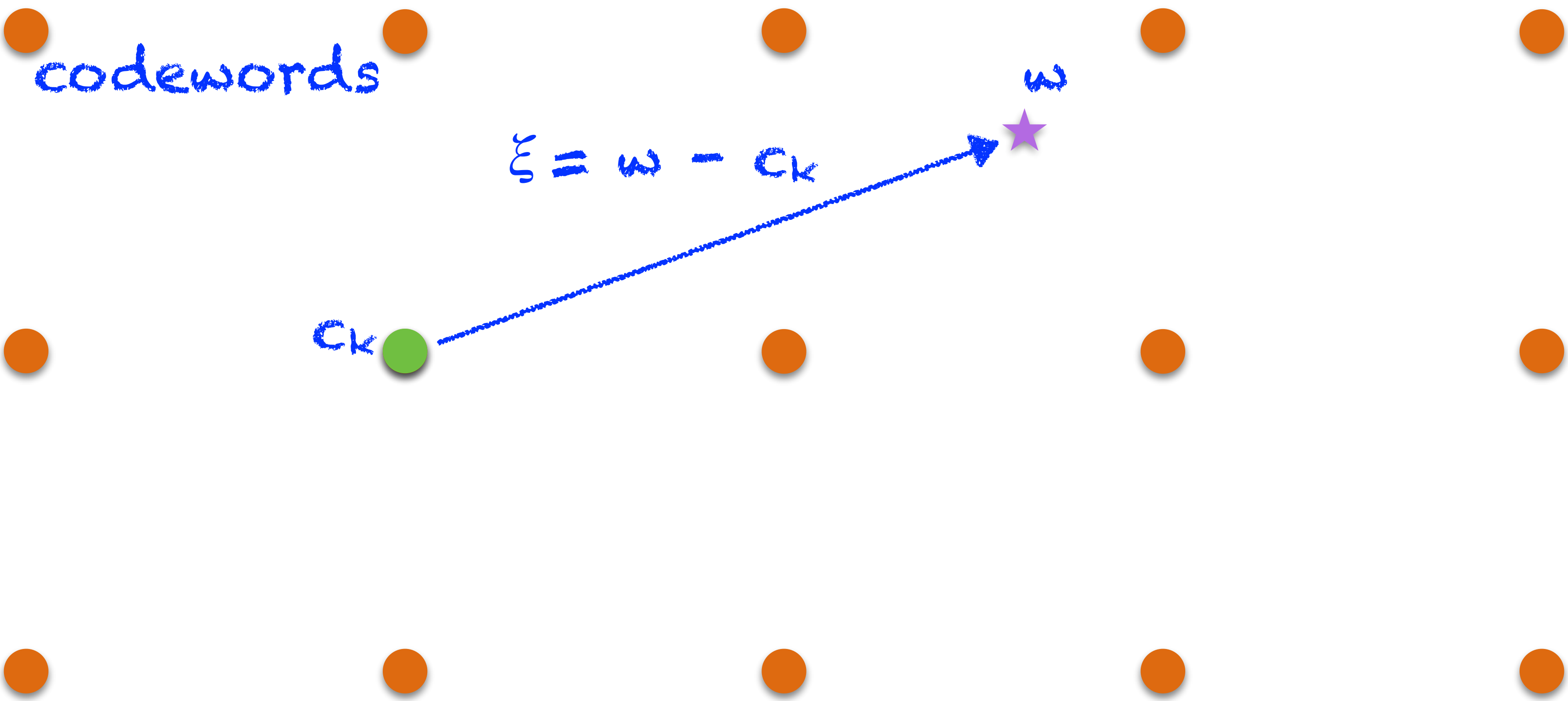
Core Principle Illustrated



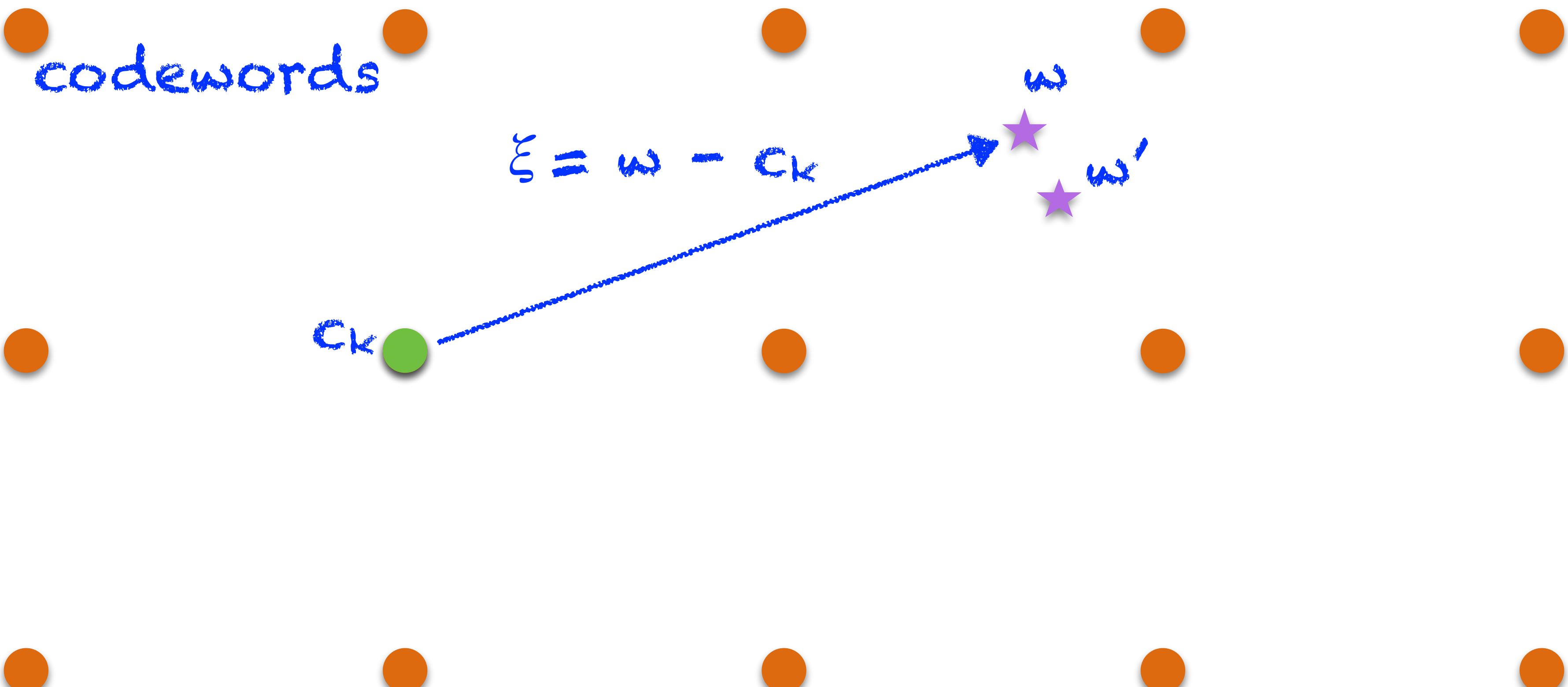
Core Principle Illustrated



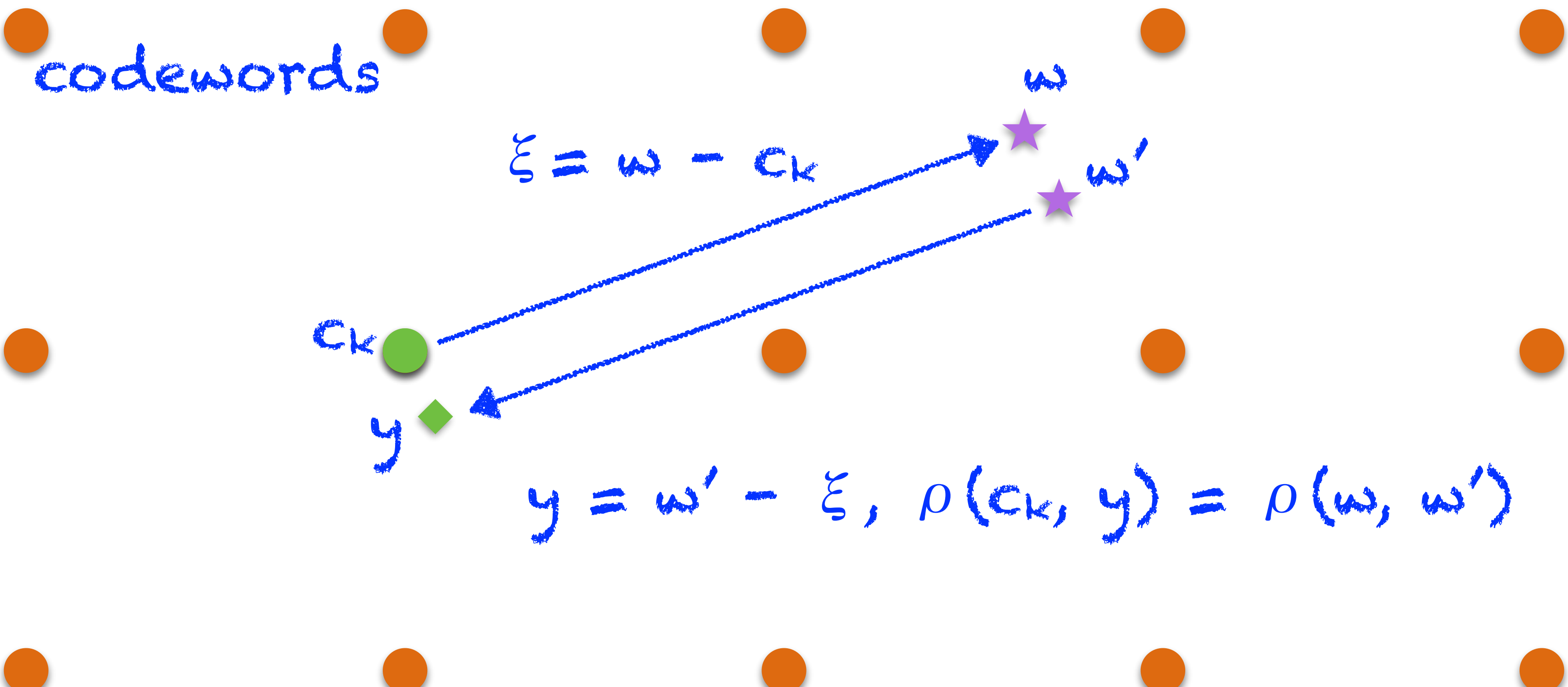
Core Principle Illustrated



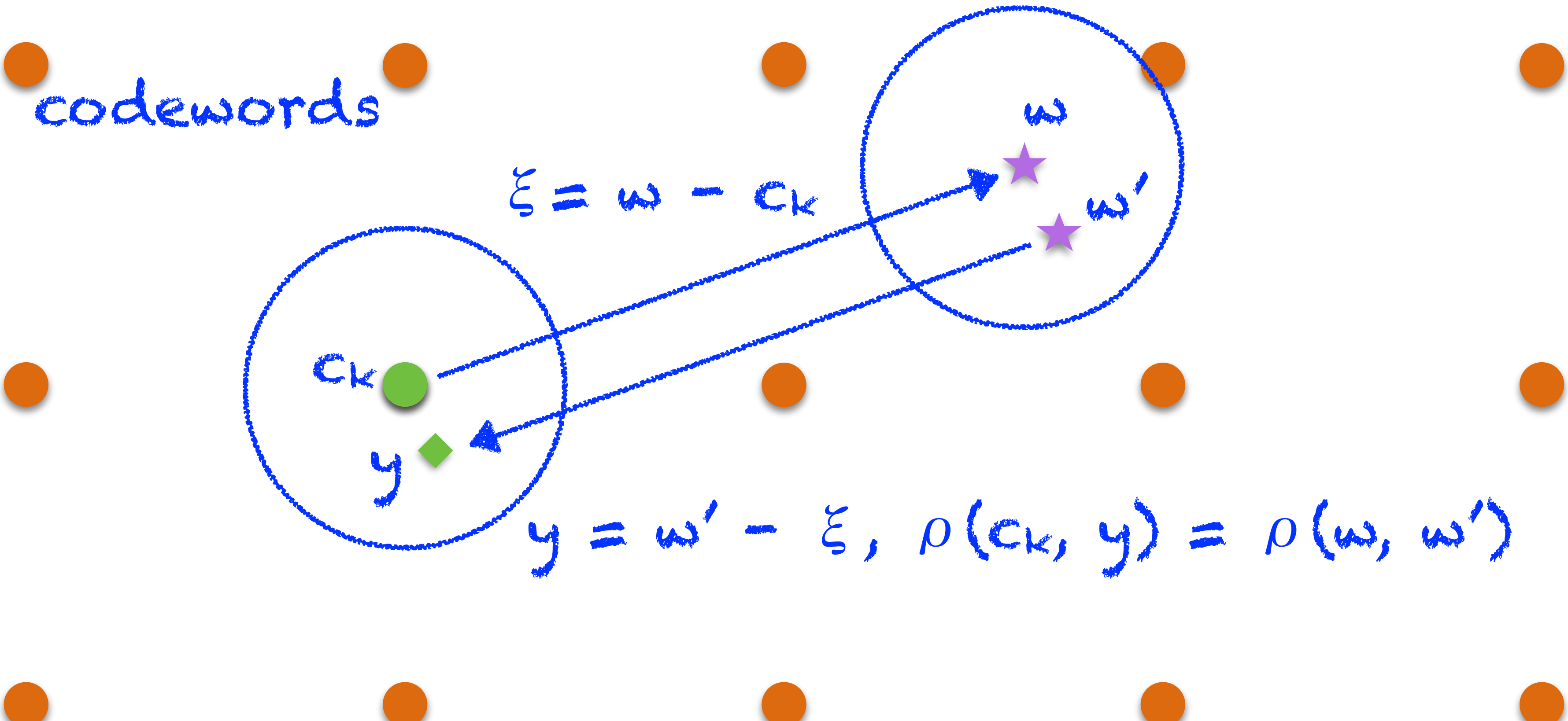
Core Principle Illustrated



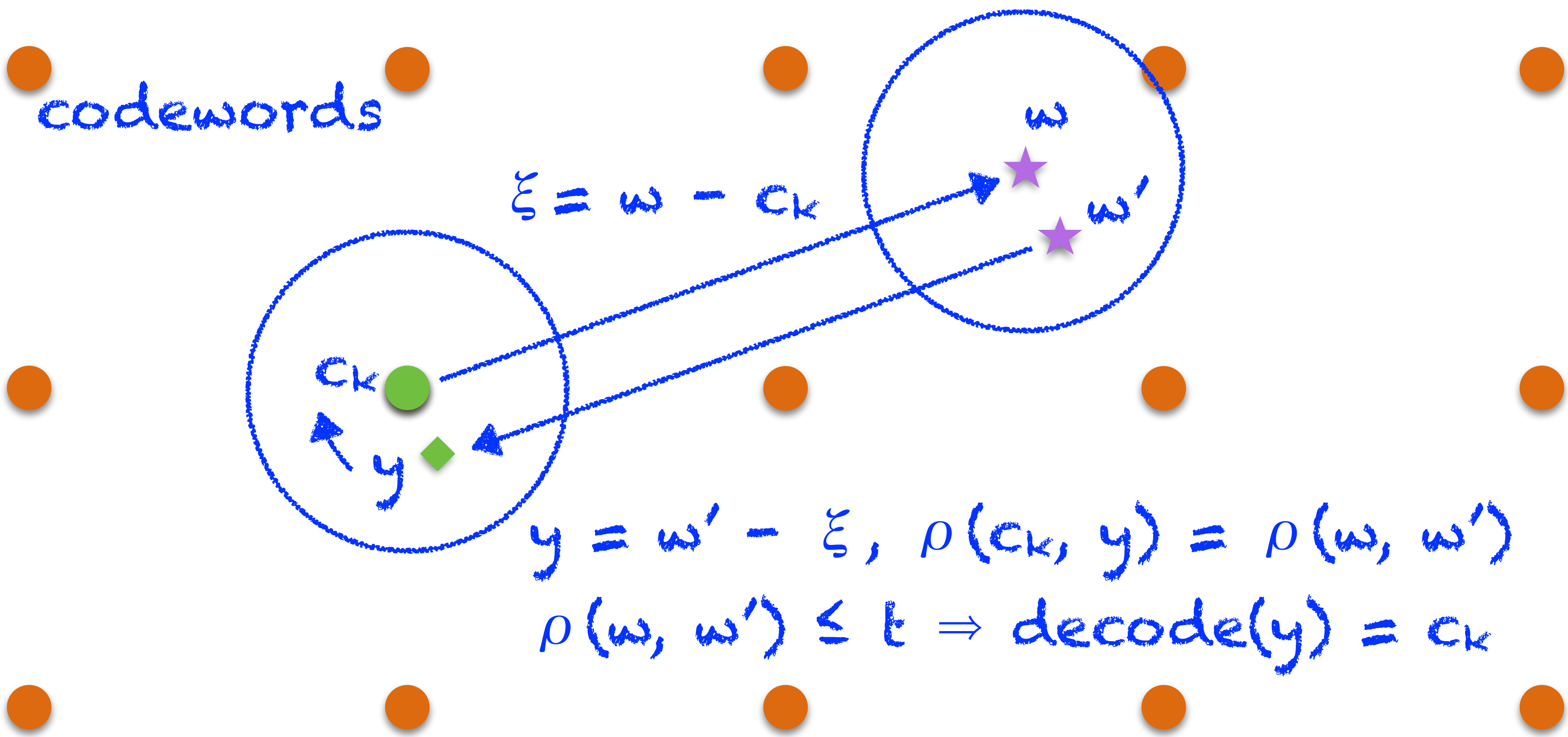
Core Principle Illustrated



Core Principle Illustrated



Core Principle Illustrated



Recovery Hint - ξ

- Let D be the redundancy of the code C in F (with respect to randomly chosen codewords).
- Having learned ξ, the attacker gets at most D bits of information on the registration BIO sample w.
  - We emphasise, we do not store any hash-print of *BIO_key* locally.
  - ξ is the only information leaked under ATA.
  - Anyway, there are schemes allowing even local template encryption under a low-entropy password, cf. below.

So, Is ξ Public?

- Unless we have a plausible algebraic model for the biometric redundancy, ξ shall not be "public" as an RSA public key, for instance.
  - We rather suggest handling it the same way as the *device_key* here.
  - Cf. also the encrypted template methods below.
- In our design, all the BIO cryptography is merely a life-saving jacket, not a silver bullet.
  - Yes, it is definitely important against ATA.
  - But we shall not overhype it!

My Voice Is My… *Entropy*

Voice-Based BIO-cryptography

- We shall start with mapping the features of the whole utterance to a *supervector w*.
- We also have to enforce an ordering such that a particular coordinate of *w* always corresponds to a particular feature variable.
  - Straightforward for text-dependent methods.
  - For text-independent methods, we can follow the trick of Baum-Welch statistics re-ordering as employed in variants of Factor Analysis by Kenny, Dehak, Brümmer, et al.

Another BIO-Crypto Protocol

- RBT ~ Randomized Biometric Templates
  - Ballard et al., 2008
  - Shares the basic idea of using an error correction mechanism to cope with intra-user variability.
  - Resulting RBT scheme can be viewed as a special kind of Fuzzy Extractor.
- Employs *randomized feature selection* together with plausible *template encryption* suitable for even a low-entropy password.

RBT Password Protection

- The authors really strived hard to devise password-based protection of the whole RBT.
  - This way, the password entropy gets combined with the BIO entropy to considerably harden ATA.
- There shall be no *verifiable plaintexts* (Lomas et al. in 1989) in RBT, so we *could* use even our precious PIN here.
  - We shall, however, verify this with respect to the particular RBT calibration we would eventually use…

Error Correction of RBT

- RBT employs a quantization of random variables for error correction.
  - This naturally introduces Euclidean distance metric for features variation.
- The role of the quantization boundary offset α_i roughly corresponds to ξ.
  - Note that α_i can be further transformed to a non-verifiable plaintext.
  - So, it can be protected by our precious PIN.

Voice-Based BKG

- BKG ~ Biometric Key Generation
  - In 2010, Carrara and Adams described a voice-based BKG by using RBT and a novel extraction of *reliable features*.
  - Euclidean metric of RBT is highly welcome here.

Text Dependency

- RBT assumes a strict order of the biometric features employed for the key derivation.
  - With the BKG based on *reliable features* extraction and RBT, this corresponds to the time order.
  - So, we get a text-dependent scheme.
- Using a feature vector derived by a variant of front-end **Factor Analysis**, we could, however, relax the time order to cover text-independent methods as well…

Recall the Joint FA Model

$$M = m + \mathbf{U}x + \mathbf{V}y + \mathbf{D}z$$

Recall the Joint FA Model

$$M = m + \mathbf{U}x + \mathbf{V}y + \mathbf{D}z$$

Speaker-specific features vector, we let $w = y$.

Another Voice-Based Scheme

- In 2001-2002, Monrose et al. employed a strict quantization together with a secret sharing scheme (SSS) to:
  - cope with intra-speaker variation,
  - allow mixing the biometric randomness with a (possibly low-entropy) password.
    - this is done via template encryption while obeying the rule of no verifiable plaintexts

Text Dependency

- To cope with ATA, the speech model part (besides the SSS) must be a speaker- and text-independent one.
  - But do not be fooled by this. This is merely to say there shall be no verifiable plaintexts (voiceprints).
  - The whole scheme, however, assumes the speaker is using the same utterance for both enrolment and key recovery.
  - ➡ Again, it is a text-dependent scheme.
  - ➡ Again, front-end **Factor Analysis** may provide us with a text-independent variant.

Towards "Back-End" Order Invariance

- There is the Fuzzy Vault scheme by Juels and Sudan since 2002.
  - Instead of SSS, they employ a noisy polynomial reconstruction based on Reed-Solomon (de)coding.
  - Furthermore, they use the quantized features directly as *x*-coordinate "probes" for the secret polynomial.
  - Finally, they employ the idea of chaffing to conceal the correct $(x, p(x))$ points.
- This scheme exhibits the important **order invariance** property, this time without front-end preprocessing tricks.
  - However, as for the VB the previous methods may be more appropriate even for TI schemes, despite the involved front-end preprocessing.

Anyway, Fuzzy Extractors Take It All

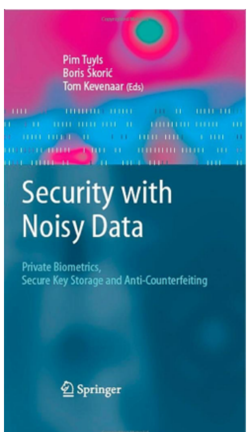
- Dodis et al. shown Fuzzy Vault can be modelled and enhanced by the general Fuzzy Extractor approach (2004).
  - Their construction is based on the set difference metric.
  - It can be seen as an improved theoretical framework for the original FV construction.
  - The idea of using a noisy polynomial reconstruction stays the same.

Too Good To Be True?

- The concise theory of Security with Noisy Data provides rather solid ground for robust protocols.
- We shall, however, verify the particular practical implementation very carefully.
  - There may be "surprisingly" new attacking strategies that were not incorporated in the former security "proofs" (Scheirer and Boult, 2007) .
  - For instance, obtaining the recovery hints for multiple enrolments of the same individual may be a problem.
  - RBT cope with this by the random feature selection.
  - Distributed implicit *BIO_key* verification also helps; **suitable entropy extractor** shall ensure *BIO_key* is decorrelated from the original biometric data (to stop spreading it)!

Conclusion

- **Fuzzy Extractors** together with the noisy data framework are the unifying theory of most of the BIO-cryptographic protocols.
  - The particular schemes developed more or less independently on FE then expose interesting practical tricks.
- To build up a real working system, we need to devise:
  - robust feature extraction,
  - error correction approach together with a suitable intra/ inter variability metric,
  - key recovery and verification scheme,
  - template protection level (with a possible entropy boost from the client password/PIN).



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tomáš Rosa, Ph.D.

Raiffeisenbank, a.s.

http://crypto.hyperlink.cz