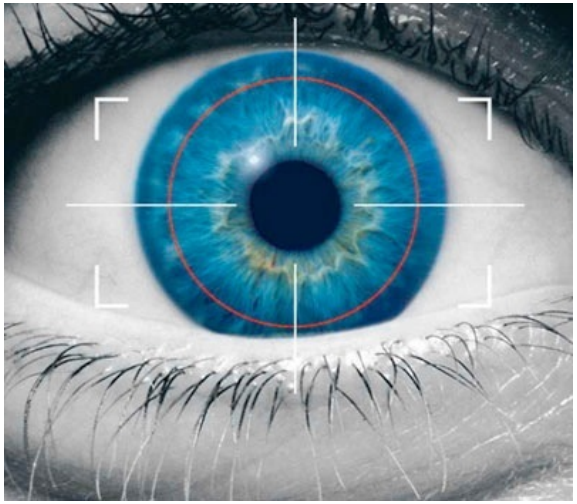
Mobile Authentication with BIO-Cryptography Taste

Tomáš Rosa
crypto.hyperlink.cz

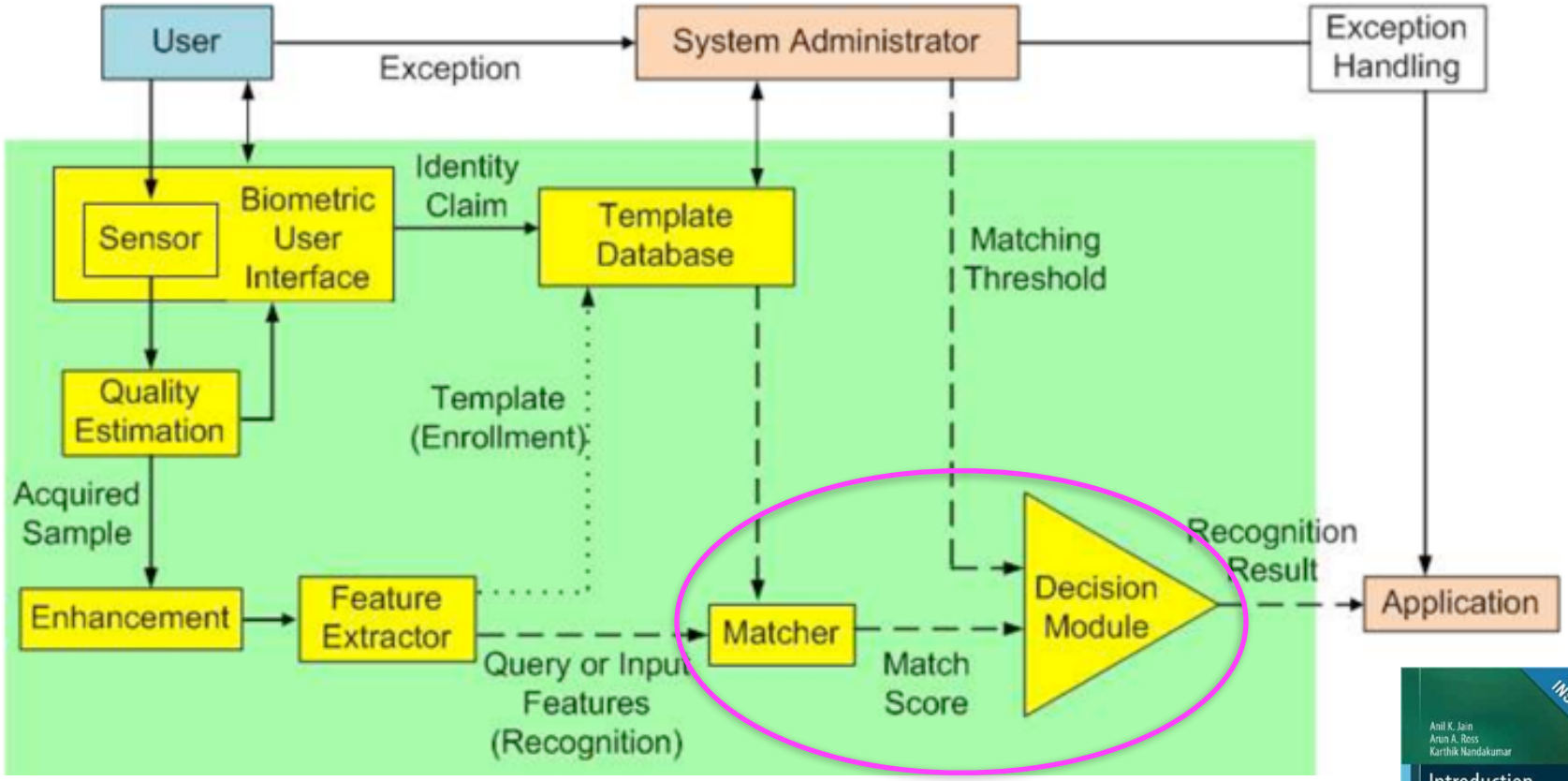
Biometric Identification/Verification

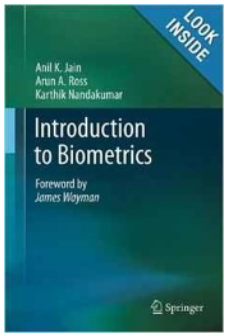
...automated establishment of the human identity based on their physical or behavioral characteristics.




Secret Files
on Biometrics

Biometric System Topology



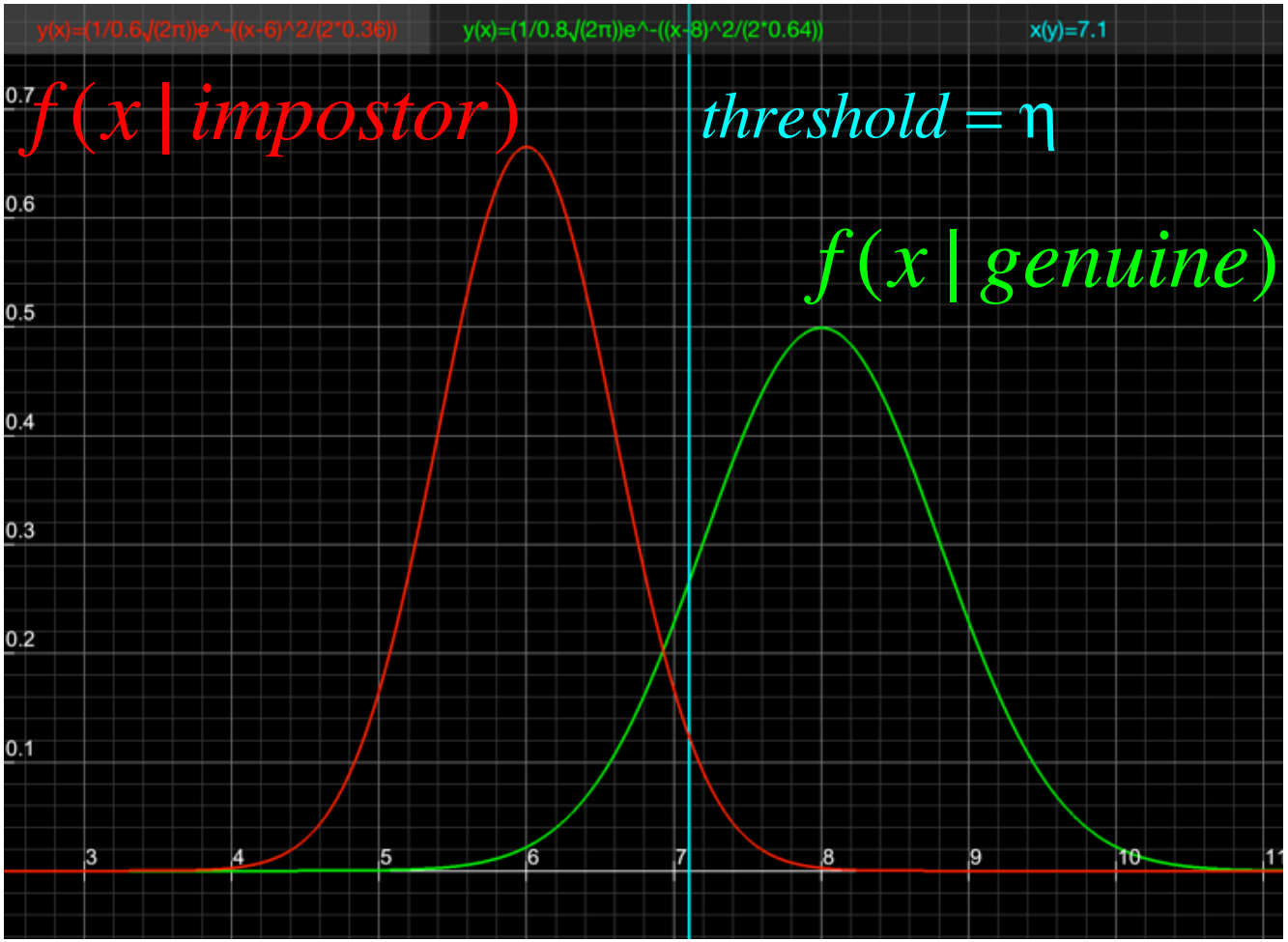


Jain, Ross, Nandakumar: Introduction to Biometrics, Springer, 2011

Match Score

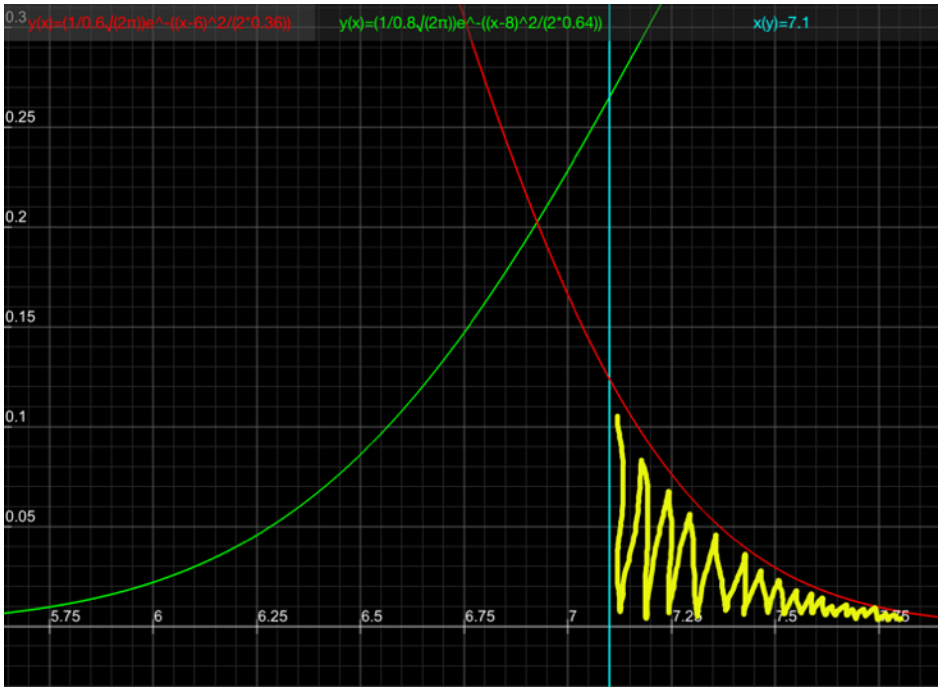
- It would be nice if we had a simple true/false result.
  - As in conventional crypto.
  - But we cannot...
- All we have is a sample of random variable *X* that follows two conditional distributions.
  - *f*(*x* | impostor)
  - *f*(*x* | genuine)

Match Score Evaluation



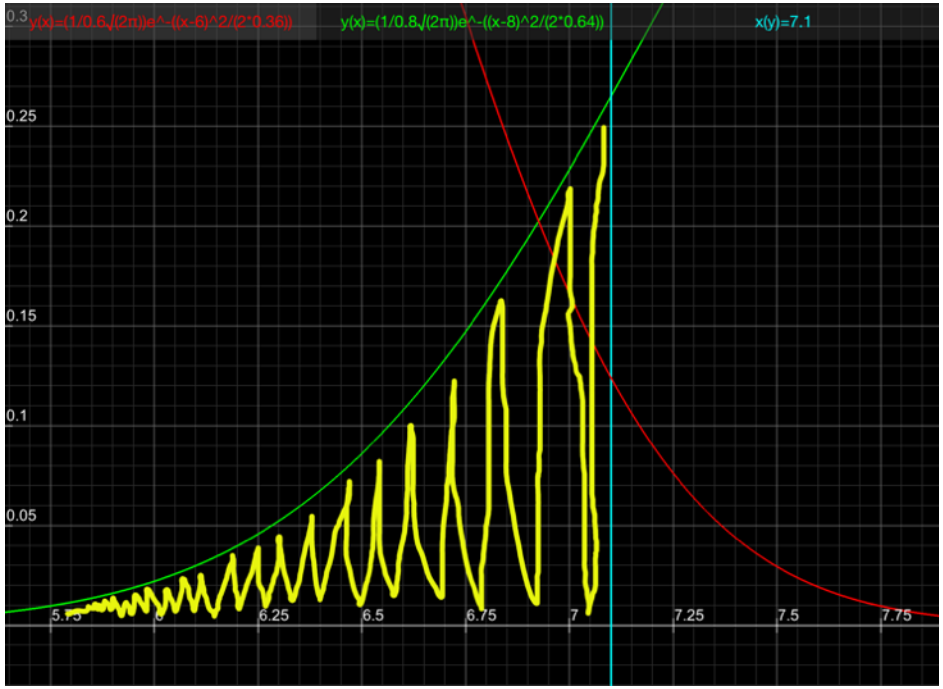
False Acceptance Rate

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

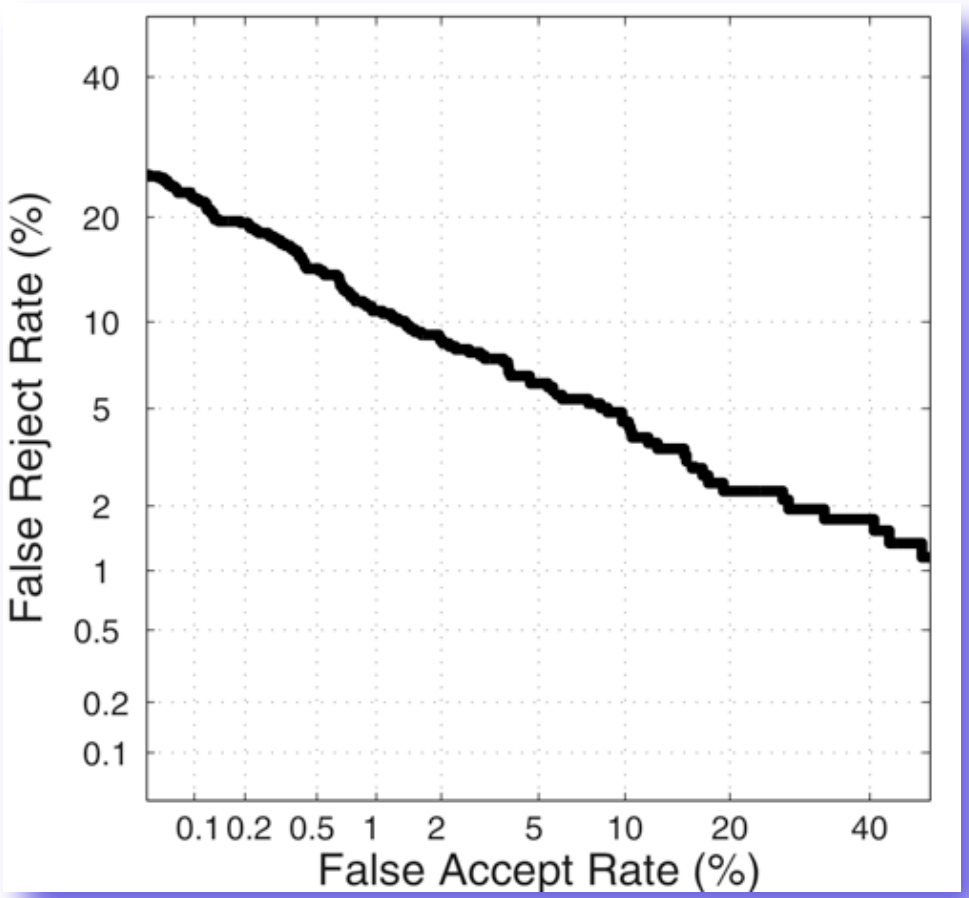


False Rejection Rate

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



Real DET Curve



Detection
Error
Tradeoff

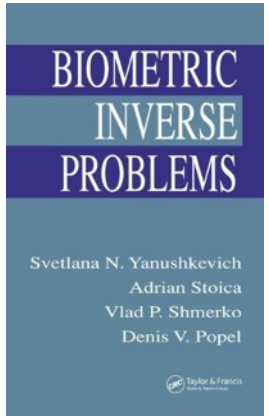
Jain, Ross, Nandakumar, Springer 2011

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

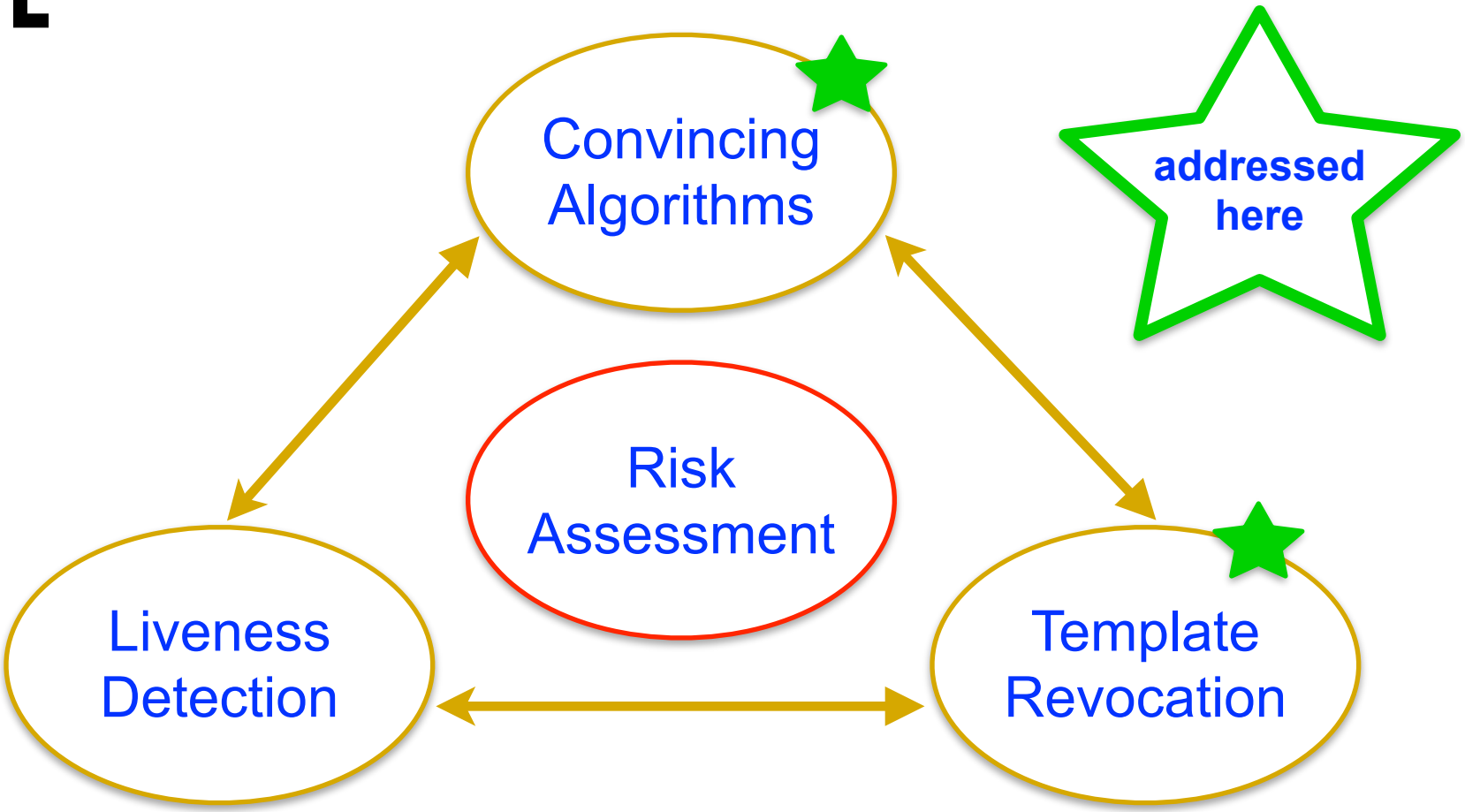
- *The process of defeating a biometric system through the introduction of fake biometric samples.*
  - *(Schuckers, Adler et al., 2010)*
- Particular modus operandi on how to deploy the attacking data vectors.
  - Can be seen as being orthogonal to the aforementioned hill-climbing attacks.

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


Do the Pentest!

Open Problems




Convincing Algorithms?


Safe Template Revocation?

Consultants Always Ready to Help!

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

Naive Approach

```
sample = get_biometric_data();

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

Naive Approach

```
sample = get_biometric_data();

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

bypassed!

Naive Approach

```
sample = get_biometric_data();

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

bypassed!

stolen!

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

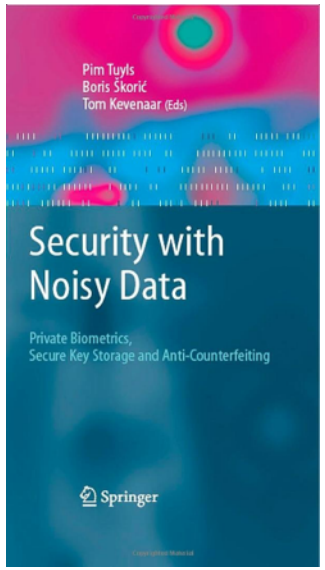
iii) let *MK* = *KDF*(*PIN_key*, *BIO_key*, *device_key*)

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

BIO_key is shared with the bank, not a BIO template

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*.



Biometric Cryptography?

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 $C \subset F, C = \{c_1, c_2, ...\}$.

$decode : F \to C$

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 ($\boldsymbol{F}$, 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).

Enrolment

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

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)

Enrolment

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

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 $w' \in \boldsymbol{F}$

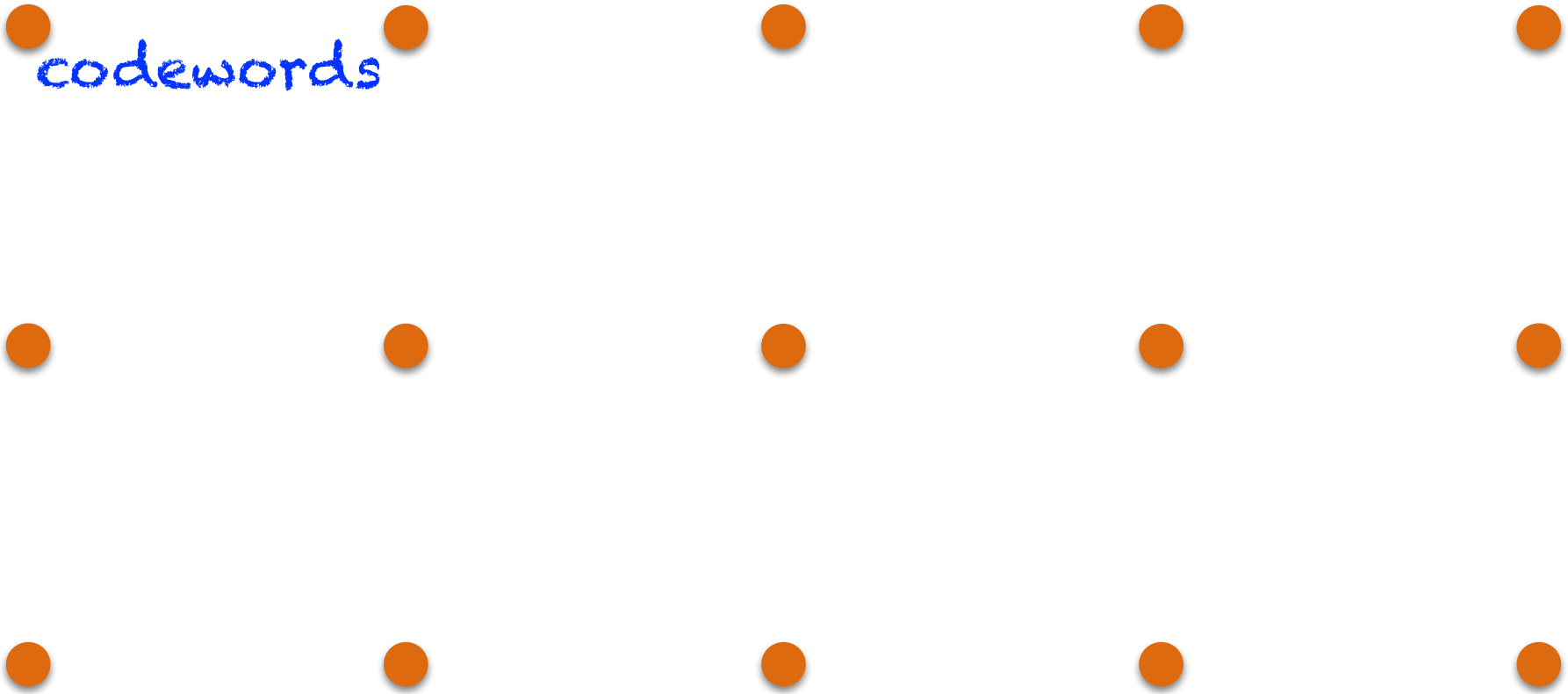
ii) let $y = w' - \xi$

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

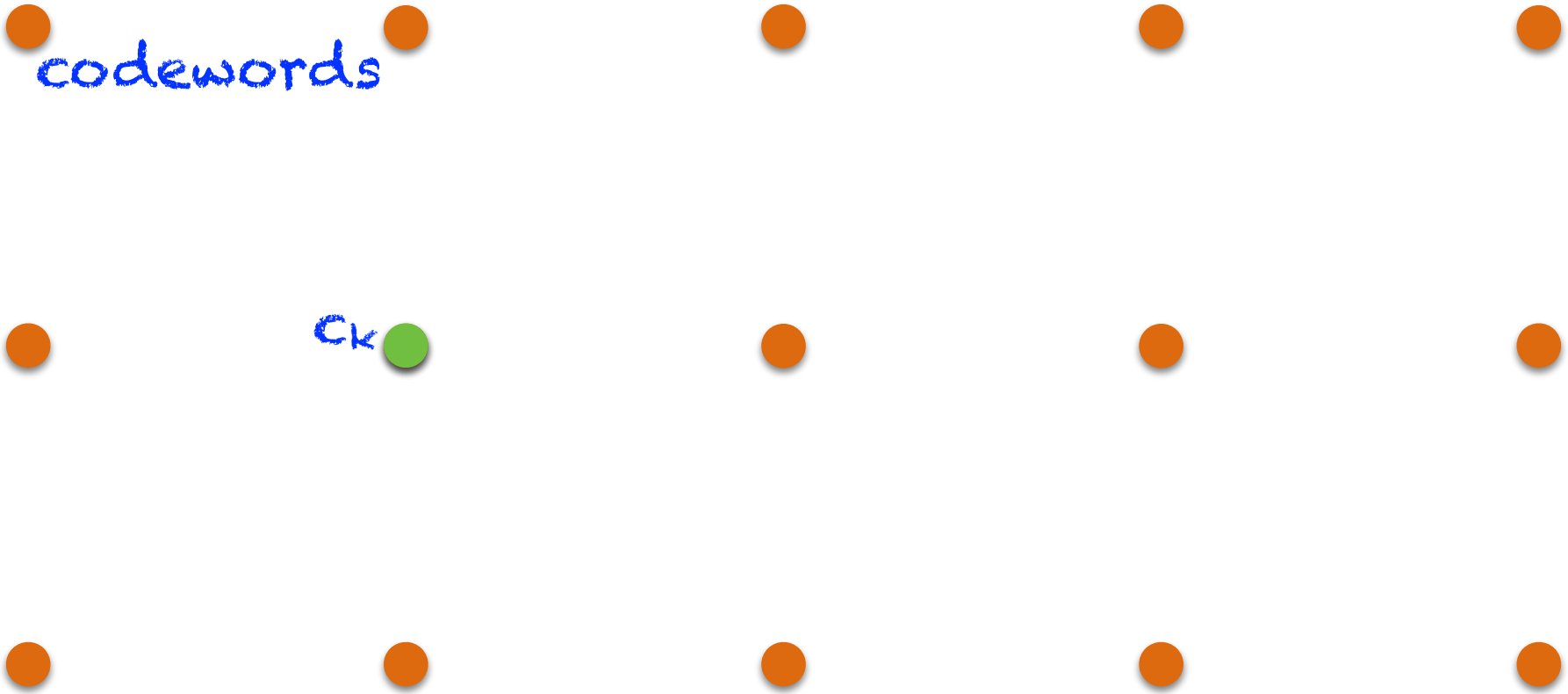
iv) let $BIO\_key' = hash(c_{key}')$

v) try to use BIO_key' in the protocol above

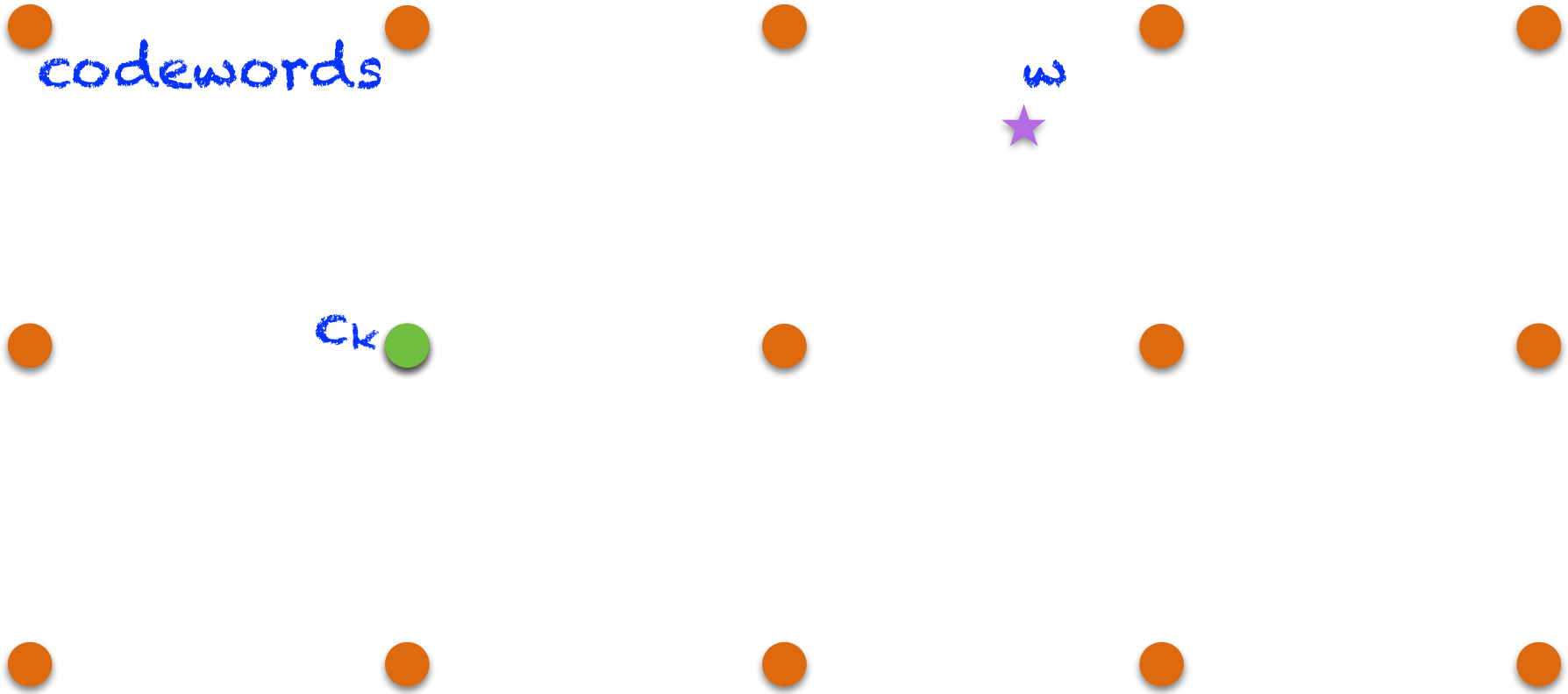
Core Principle Illustrated



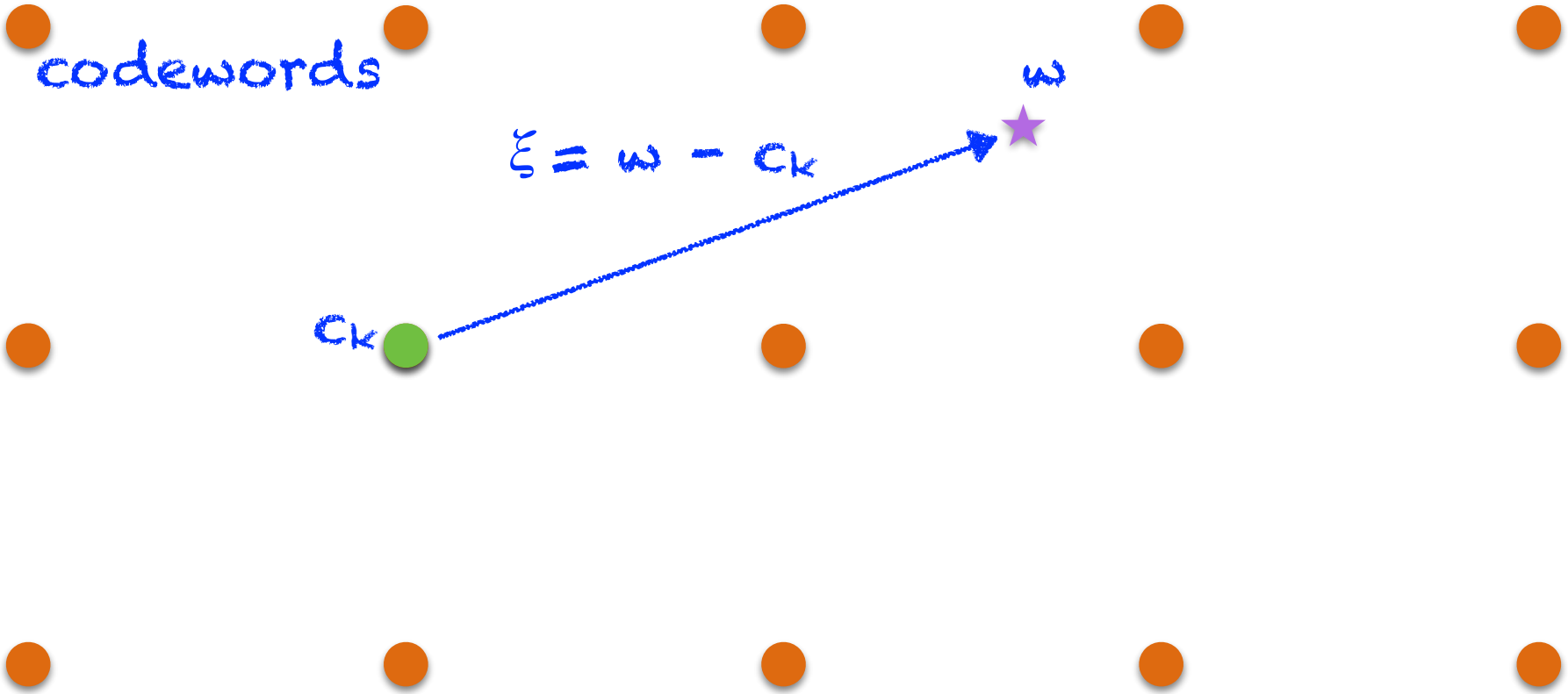
Core Principle Illustrated



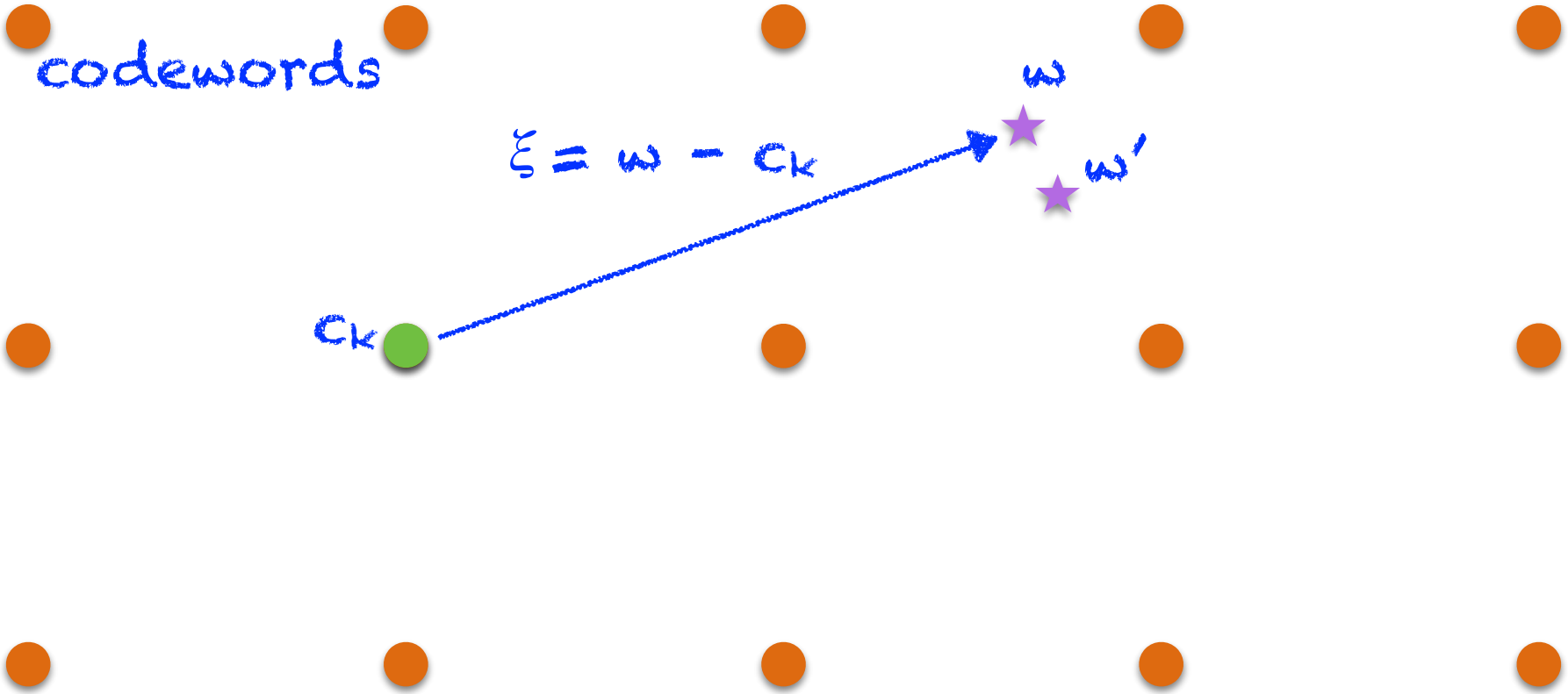
Core Principle Illustrated



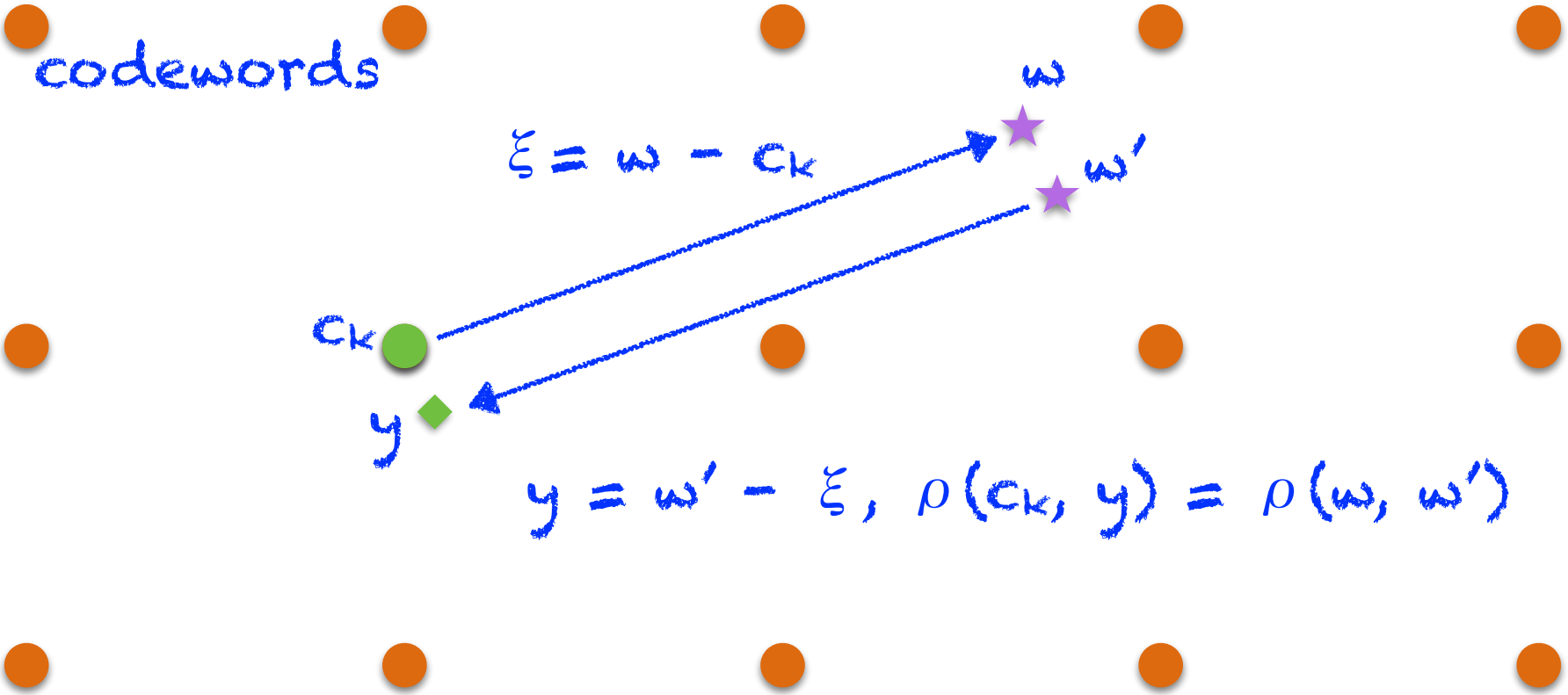
Core Principle Illustrated



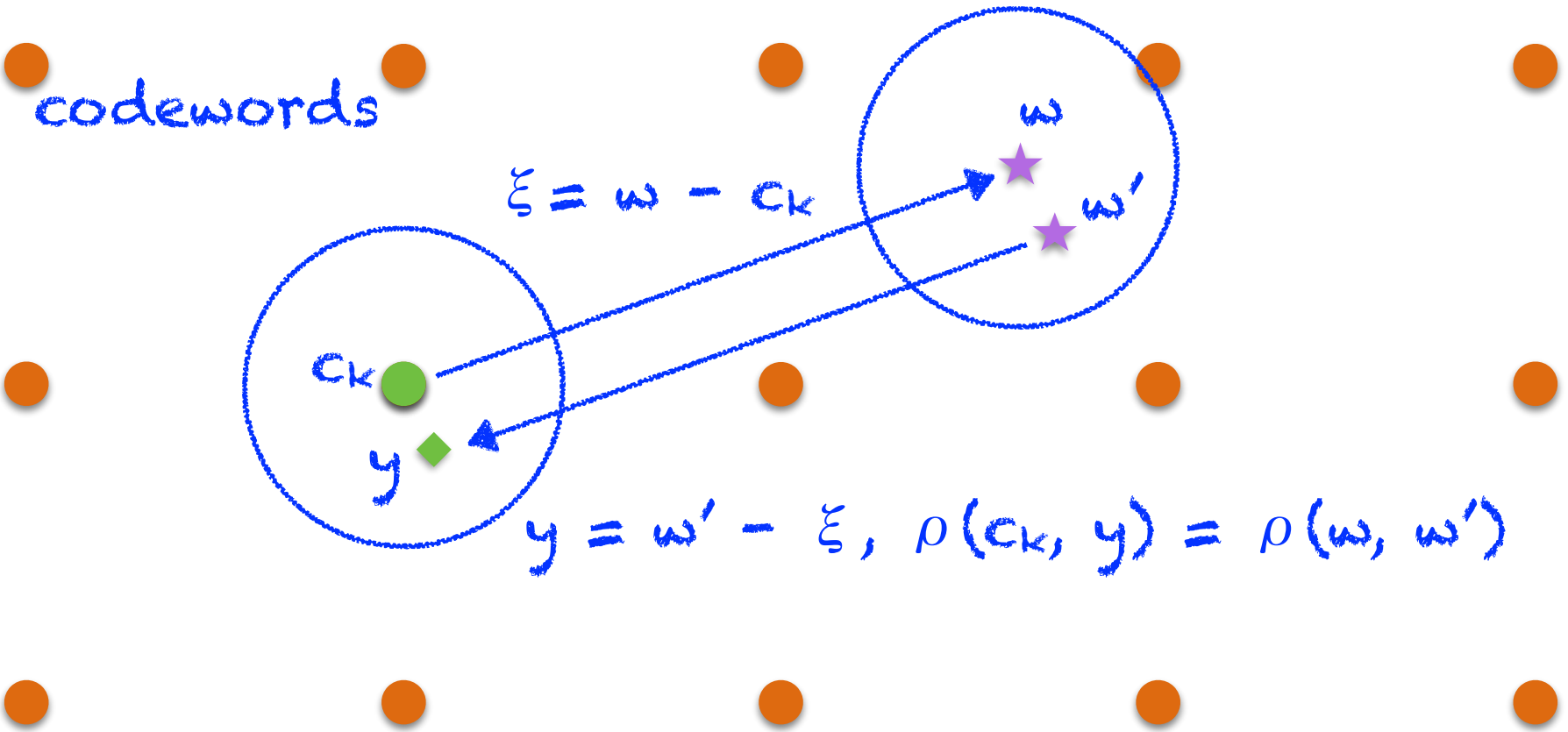
Core Principle Illustrated



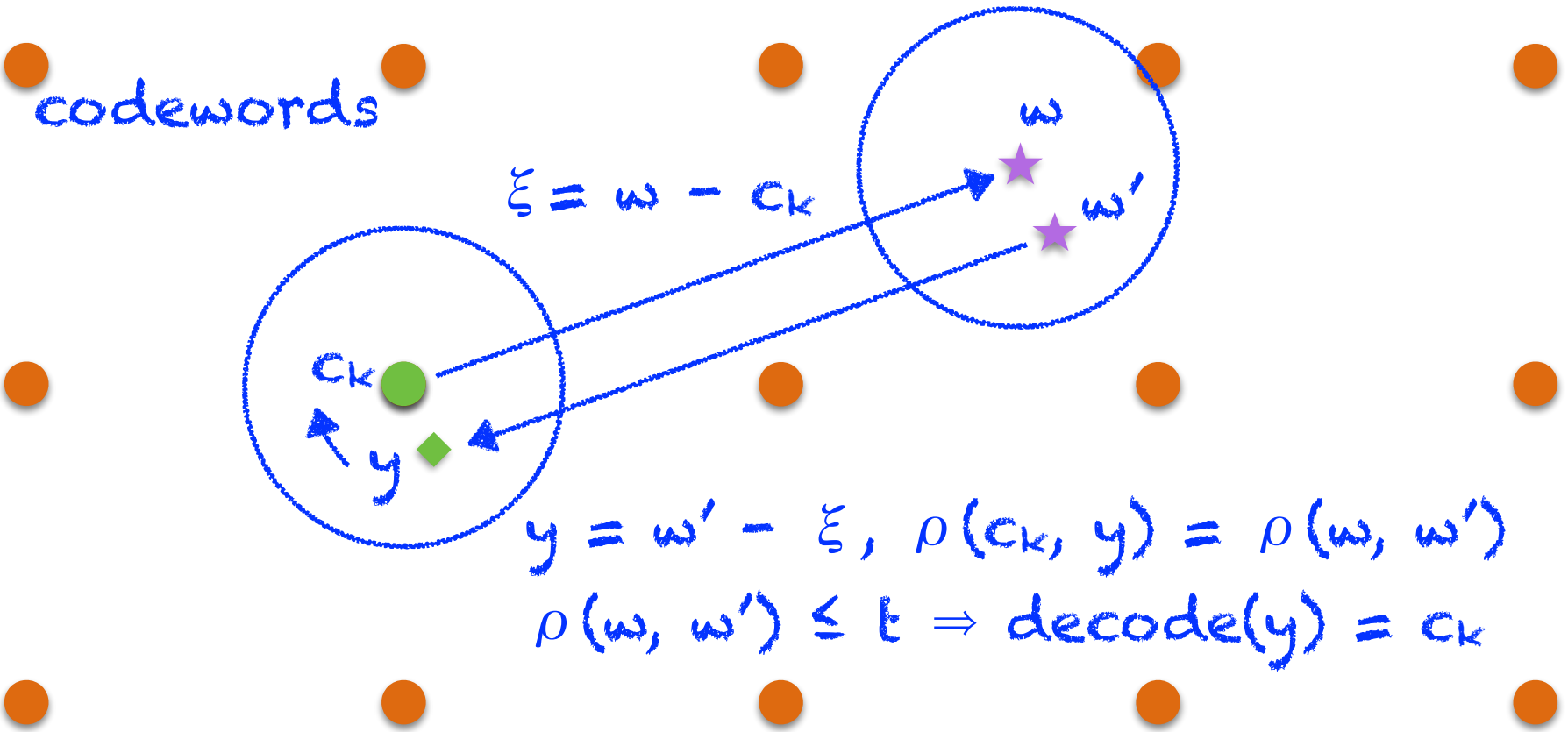
Core Principle Illustrated



Core Principle Illustrated



Core Principle Illustrated



Recovery Hint - ξ

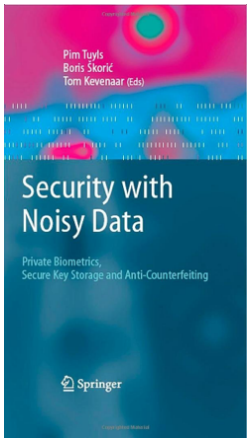
- Let *D* be the redundancy of the code *C* in *F* (with respect to randomly chosen codewords).
- Having learned ξ, the attacker gets at most *D* bits of information on the registration BIO sample *w*.
  - We emphasise, we do not store any hash-print of *BIO_key* locally.
  - ξ is the only information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

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.
- In our design, all the BIO cryptography is merely a life-saving jacket, not a silver bullet.
  - Yes, it is definitely important against a physical device theft.
  - But we shall not overhype 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).



Thank You For Attention

Tomáš Rosa, Ph.D.

http://crypto.hyperlink.cz

Movie Snapshots Taken From

- *Tajemství hradu v Karpatech*, ČR, 1978
- *Císařův pekař*, ČR, 1951
- *The Magnificent Seven*, United Artists, USA, 1960
- *Slunce, seno, jahody*, ČR, 1983