Spoofing and deepfakes in biometrics and the risks of identity theft

AI Tech & Policy Talks

Prof. Sébastien Marcel (www.idiap.ch/∼marcel)

Nov 15th 2022

Idiap Research Institute

Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

The BSP group at Idiap

Research themes

signal (image, audio) processing and AI (machine/deep learning) applied to BSP:

- Biometric and Attribute recognition:
  - **Face** (2D, 3D, multi-spectral, heterogeneous)
  - **Speaker**, **Vein** (finger, palm and wrist)
  - Electro-physiology (EEG/ECG)
  - Gender recognition, age and heart-beat estimation (**rPPG**)
- Security: Presentation Attacks (PA aka **spoofing**), Morphing and Deepfakes detection
- Privacy: Template Protection (**irreversible** and **unlinkable** biometrics)
- Multi-modal fusion: combining biometrics and PA detection (PAD)
- AI and responsible datasets: **fairness**, trojan/backdoors, ethics and **synthetic datasets**

Reproducible Research as a priority

Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

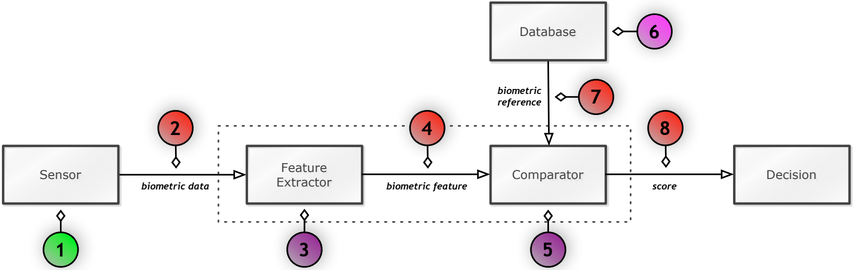
PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

Biometric system vulnerabilities

- A biometric system is vulnerable to many types of **attacks**[1]



- We are mostly interested in attacks on the sensor (1), referred to as **Presentation Attacks**
- We will also consider **Presentation Attack Detection**

[1] Ratha, N. K. *et al.* "Enhancing security and privacy in biometrics-based authentication systems", *IBM Systems Journal*, 40(3), pp. 614–634 (2001)

Definitions

Presentation Attack (PA)

- An attempt to **fool** the biometric recognition system by presenting **fake** biometric data to the sensor, e.g.,
    - A **replica** of an enrolled user's biometric features (if the goal is to **impersonate** that user), or
    - **Generic** biometric features (if the goal is to **avoid recognition**)
- PAs are also commonly called **spoofing attacks**, and the fake biometric data is referred to as a **spoof**

Presentation Attack Detection (PAD)

- The **determination** of a PA (i.e., "the presented biometric data is/is not a spoof")
- Also commonly referred to as **anti-spoofing**

Definitions

Presentation Attack Instrument (PAI)

- The **biometric characteristic** or **object** used to launch a PA
- *Examples:* Face mask, gummy fingerprint, dead body parts, etc.

Bona Fide Presentation

- **Normal** (intended) interaction of the subject with the biometric system's sensor
- Basically, anything which is **not** a PA

Note: See *Biometric presentation attack detection – part 1*, ISO/IEC 30107-1:2016 (2016) for formal (standardised) definitions.

Importance

- PAs pose a **major threat** to biometric recognition systems
- This is because the attack is **external** to the system (i.e., at the sensor), so the attacker does **not** need to have any **knowledge** about the **internal workings** of the system
- In fact, PAs can be launched by basically **anyone**, often using very **basic tools**
- Let's look at a few examples of PAs in **reality** (real-life attempts at using PA to either conceal or steal identities)

Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

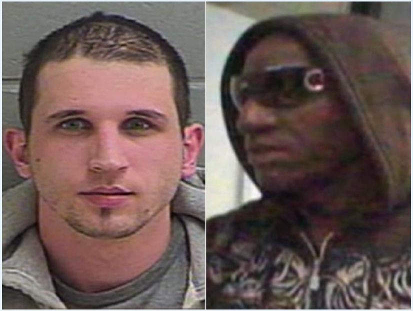
PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

Face

Robbery (2010)

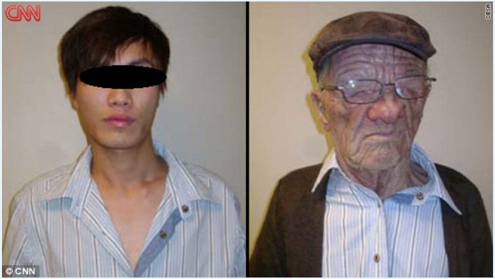


Conrad Zdzierak used a silicone face mask to pass himself off as a black character "SPFX The Player" during bank robberies[2]

[2] http://www.telegraph.co.uk/news/worldnews/northamerica/usa/8193185/US-criminals-using-film-quality-masks-during-bank-robberies.html

Face

Immigration (2011)



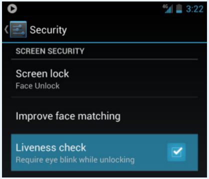
A young Asian man disguised himself as an old Caucasian man using a silicone face mask, boarded a plane in Hong Kong, then removed the disguise mid-flight and asked for refugee status upon arriving in Canada[3]

[3] http://www.dailymail.co.uk/news/article-1326885/Man-boards-plane-disguised-old-man-arrested-arrival-Canada.html

Face

Smartphone unlock (2011)

- The Face Unlock feature on Galaxy Nexus, running Android 4.0, was spoofed by a face photograph[4]
- Android 4.1 added a “liveness check” (eye blink)



Security

SCREEN SECURITY

Screen lock
Face Unlock

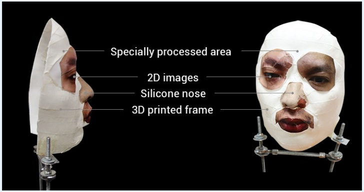
Improve face matching

Liveness check ☑
Require eye blink while unlocking

[4] http://www.geek.com/android/android-face-lock-feature-spoofed-by-photograph-1440953

Face

Smartphone unlock (2017)

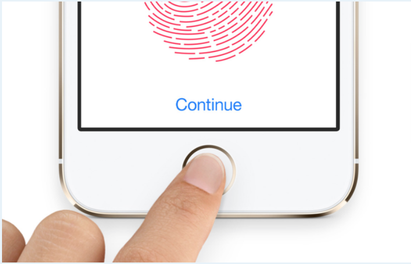


iPhone X's Face ID was spoofed by a specially crafted face mask[5], despite claims that it is robust to mask attacks

[5] https://www.youtube.com/watch?v=i4YQRLQVixM

Fingerprints

Smartphone unlock (2013)



The fingerprint unlock feature (Touch ID) of iPhone 5s was spoofed using a fake fingerprint[6]

[6] https://www.youtube.com/watch?v=HM8b8d8kSNQ

Iris

Smartphone unlock (2017)

The iris unlock feature in Samsung Galaxy S8 was spoofed using a printed photograph of the enrolled user's iris (plus a contact lens)[7]

[7] https://www.youtube.com/watch?v=gtQ4yzbsi-c

Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

PAI

- Recall that a Presentation Attack Instrument (PAI) is the **biometric characteristic** or **object** used to launch a PA
- We can thus think of a PAI as a **method** for **executing** a PA
- The previous PA examples demonstrated a few different types of PAIs (e.g., face masks, printed finger vein patterns, etc.)
- We will now consider PAIs in more detail, with a specific focus on **face** PAIs (since this is the main area of expertise for Idiap's Biometrics Security and Privacy group)
- Feel free to try these PAIs at home, but not for committing crimes!

Types of PAIs

- In general, PAIs can be divided into 3 main categories:
  1. **Photograph or recording:** The attacker acquires a photograph or recording of the target's biometric characteristic, and replays it to the biometric recognition system
  2. **Synthetic biometric characteristic:** The attacker generates a synthetic model of either the target's biometric characteristic or a generic feature set, then presents it to the biometric recognition system
  3. **Self-modification:** The attacker alters, physically or digitally, their own biometric characteristic(s) to either mimic the target's biometric characteristic(s) or to simply avoid being recognised as themselves
- Let's explore each of these PAI categories in turn, with a few examples for each

PAI: Photograph or recording

Printed face image

- The method:
  1. Print an image of the target's face
  2. Present the face image to the face recognition system

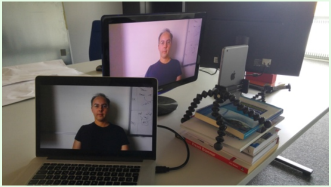
- This method was demonstrated[8] to work in launching a PA on 3 commercial face recognition systems

[8] Nguyen, D. *et al.* "Your Face Is NOT Your Password", *Black Hat* (2009)

PAI: Photograph or recording

Digital face image or video

- The method:
  1. Capture a digital image or record a video of the target's face (e.g., using a smartphone or tablet)
  2. Present the image or video to the face recognition system

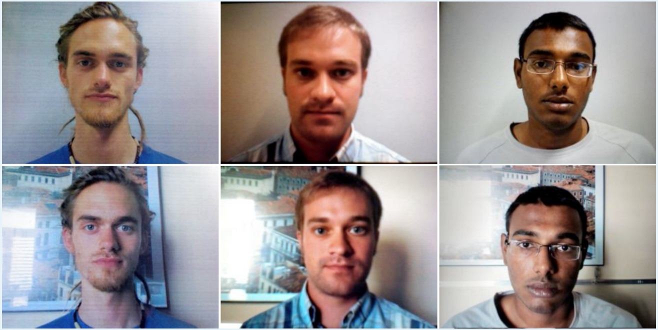


- The REPLAY-MOBILE database[9] contains face PAIs consisting of digital (as well as printed) photographs and videos acquired in various scenarios

[9] Costa-Pazo, A. *et al.* "The REPLAY-MOBILE Face Presentation-Attack Database", *IEEE BIOSIG* (2016)

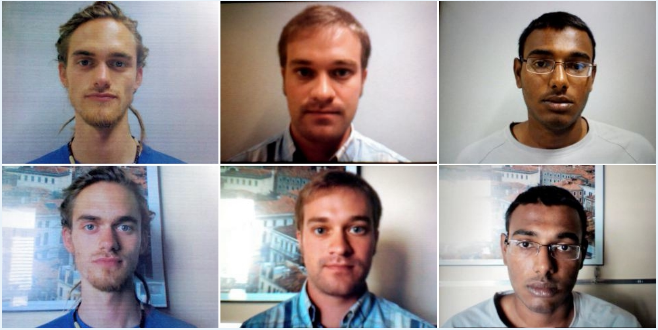
Quizz time !

Which images are Bona Fide and which are PA?



Quizz time !

Which images are Bona Fide and which are PA?



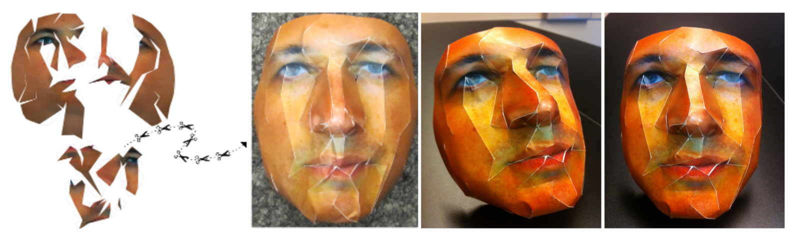
All are PAs!

- *Left:* Printed images
- *Middle:* iPhone (digital) images
- *Right:* iPad (digital) images

PAI: Synthetic biometric characteristic

Paper face mask

- The method:
  1. Acquire 1 frontal and 2 profile 2D images of the target's face
  2. Upload the images to ThatsMyFace.com, where a 3D model of the target's face will be generated and the corresponding paper net will be mailed to you
  3. Craft the net to construct a 3D paper mask of the target's face

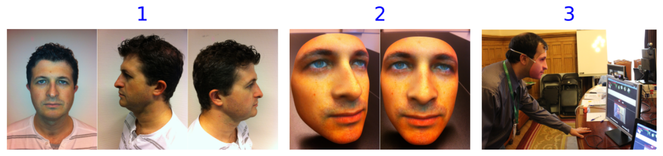


- Cheap to make ($\approx$ 25 USD), but not very effective

PAI: Synthetic biometric characteristic

Hard (resin composite) face mask

- The method[10]
  1. Acquire 1 frontal and 2 profile 2D images of the target's face
  2. Upload the images to `ThatsMyFace.com`, where a 3D model of the target's face will be generated and the corresponding hard face mask (made of a resin composite) will be mailed to you
  3. Present the mask to the face recognition system



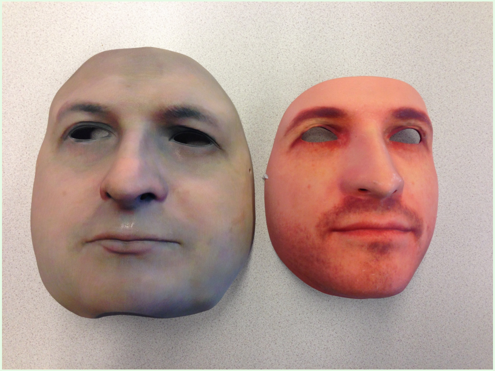
- More expensive ($\approx$ 300 USD) than paper masks, but better
- Vivid colours, but no natural face movement

[10] Erdogmus, N. and Marcel, S. "Spoofing Face Recognition with 3D Masks", *IEEE TIFS*, 9(7), pp. 1084–1097 (2014)

PAI: Synthetic biometric characteristic

Hard (resin composite) face mask with eye holes

- Same as previous example, except this time the mask has eye holes

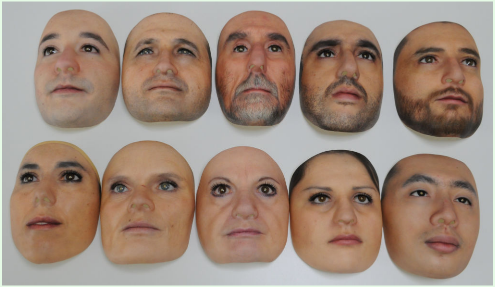


- Allows for eye movement, so a little more natural, but no flexibility for facial movement

PAI: Synthetic biometric characteristic

Hyper-realistic face masks

- Same as previous example but hyper-realistic[11] using HiRes pictures



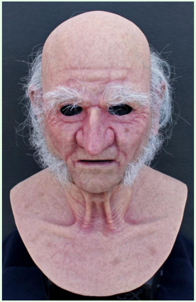
- More expensive (≈ 3,000 USD) per mask

[11] http://real-f.jp

PAI: Synthetic biometric characteristic

Silicone face mask – generic

- A generic silicone face mask could be used to obfuscate an attacker's identity, but it does not correspond to any specific target face



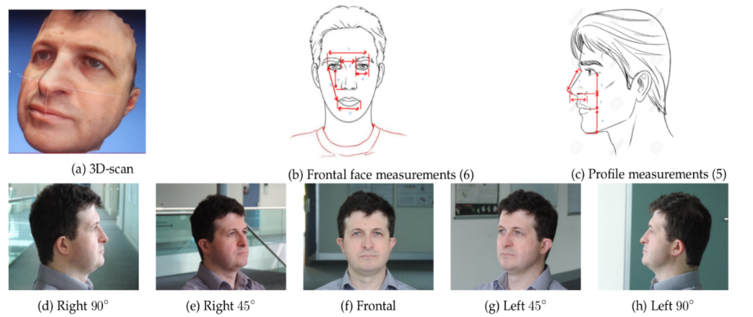
- Generic silicone face masks can be bought from a manufacturer such as CFX[12], for ≈ 800 USD

[12] https://www.compositeeffects.com

PAI: Synthetic biometric characteristic

Silicone face mask – customised

- The method[13]
  1. Acquire a 3D scan, measurements, and multiple 2D colour images of the target's face



(a) 3D-scan

(b) Frontal face measurements (6)

(c) Profile measurements (5)

(d) Right 90°

(e) Right 45°

(f) Frontal

(g) Left 45°

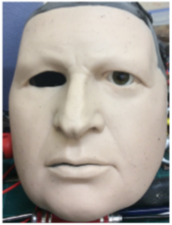
(h) Left 90°

[13] Kotwal, K. *et al.* "Multispectral Deep Embeddings As a Countermeasure To Custom Silicone Mask Presentation Attacks", *IEEE T-BIOM*, 4(1), pp. 238–251 (2019)

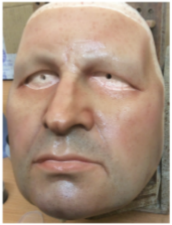
PAI: Synthetic biometric characteristic

Silicone face mask – customised

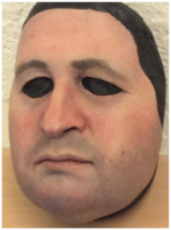
2. Send the information to a manufacturer (e.g., Nimba Creations[14]), who will generate a customised 3D silicone mask, including manual application of facial features (e.g., skin colour, eyebrows, etc.), for ≈ 4,000 USD



Raw mask



Intermediate mask



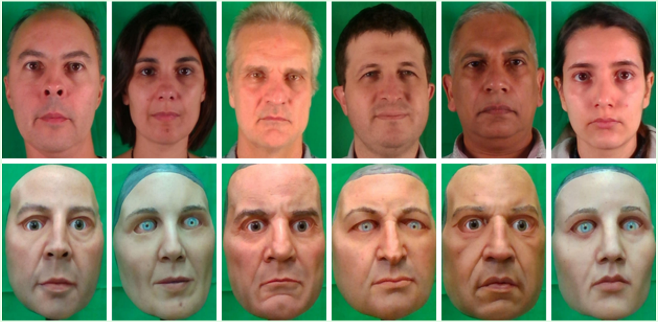
Final mask

3. Present the mask to the face recognition system

[14] https://www.nimbacreations.com/

PAI: Synthetic biometric characteristic

Silicone face mask – customised



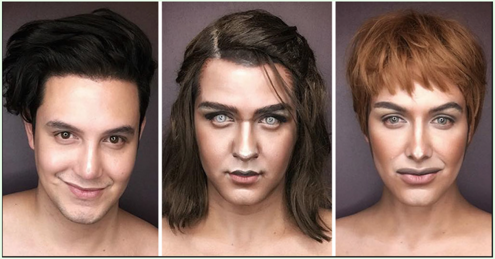
- The customised silicone face masks are quite life-like and they allow for some flexibility in facial movement
- Effective for launching PAs against face recognition systems[15]

[15] Ramachandra, R. *et al.* "Custom silicone Face Masks: Vulnerability of Commercial Face Recognition Systems & Presentation Attack Detection", *IEEE IWBF*, pp. 1–6 (2019)

PAI: Self-modification

Face make-up

- Apply make-up to the attacker's face to impersonate an enrolled user of a face recognition system:

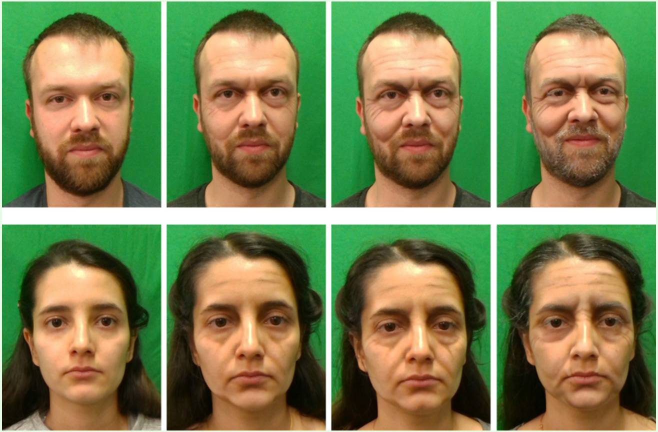


- PAs look realistic and allow for natural facial motion

PAI: Self-modification

Face make-up

- Apply make-up to the attacker's face to avoid being recognised, e.g.:

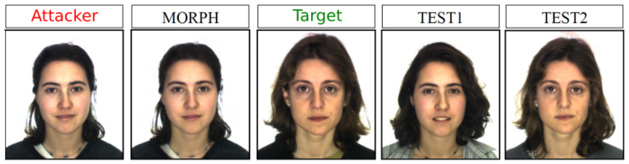


- PAs look realistic and allow for natural facial motion

PAI: Self-modification

Morphed face image

- The method:
  1. Digitally morph the attacker's face image by combining it with the target's face image
  2. Enroll the morphed image into the face recognition system
  3. Present the attacker's *or* the target's face to the face recognition system – both should match the morphed image!



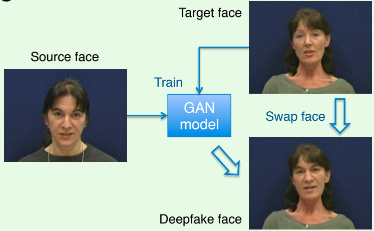
- This method was demonstrated[16] to work in launching a PA on two face recognition systems
- Potential problem for biometric passports

[16] Ferrara, M. *et al.* "The Magic Passport", *IEEE/IAPR IJCB* (2014)

PAI: Self-modification

DeepFakes

- The method:
  1. Create a video of the attacker saying or doing something
  2. Map the target's face onto the attacker's face



e.g., https://www.youtube.com/watch?v=3f66kBwfMto

- Usually accomplished using a type of deep-learning network called a Generative Adversarial Network (GAN)
- DeepFakes are becoming increasingly more sophisticated and have the potential to become a real problem (“fake news”)

Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

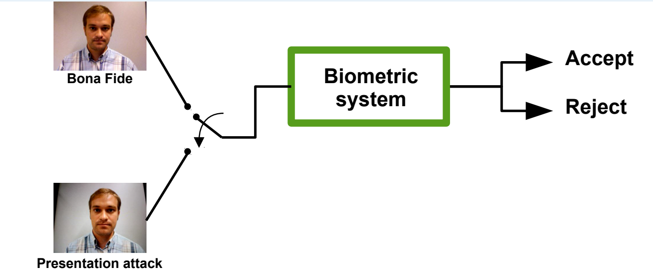
PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

Biometrics and PAD

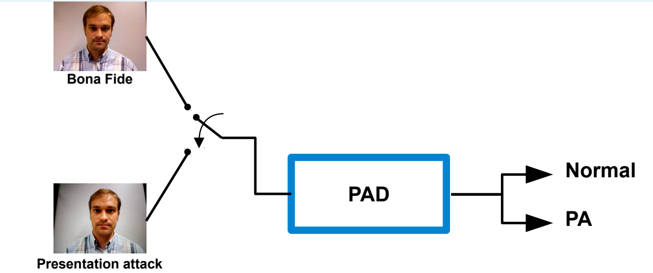
Biometric sub-system: a binary classifier



biometric recognition just compares a probe to a reference and measures "similarity" to accept genuine users and to reject impostors.

Biometrics and PAD

PAD sub-system: a binary classifier

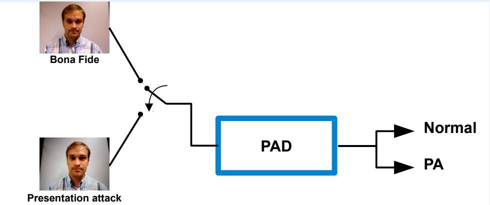


PAD determines if the probe is bona fide or a PA.

Biometrics and PAD

PAD methods

- software-based: biometric data from the sensor is analysed to discriminate bona fide vs PA (eg. motion, texture)
- hardware-based: an additional sensor (eg. multi-spectra) is used and its data analysed to discriminate bona fide vs PA (eg. heart beat, 3D, thermal imaging, multi-spectral e.g. NIR/SWIR)
- challenge-response: the user interacts with the system (eg. prompted text in face/speaker recognition)



Outline

Biometrics Security and Privacy (BSP) at Idiap

Introduction to PAD

PA examples – reality

PA Instruments (PAIs)

Presentation Attack Detection (PAD)

Conclusion

Conclusion

Is PAD a solved problem?

- Biometrics is more prevalent hence **incentives** for launching PAs are multiplying
- PAD solutions are deployed but proper certification is lacking
- Active PAD research but generalisation (to unseen attacks) is challenging – **arms race**
- PAD is not a completely solved, it continues to be an **important field of research**

Reference

- Handbook of Biometric Anti-Spoofing (Ed 2), S. Marcel and al. (2019)

Thank you for your attention!

Prof. Sébastien Marcel (www.idiap.ch/∼marcel)

Idiap Research Institute, Martigny, Switzerland