Remote KYC: Attacks and Counter-Measures

Marc PIC
Digital Labs
SURYS
Bussy-Saint-Georges, France
m.pic@surys.com

Gaël MAHFOUDI
Digital Labs
SURYS
Bussy-Saint-Georges, France
g.mahfoudi@surys.com

Anis TRABELSI
Digital Labs
SURYS
Bussy-Saint-Georges, France
a.trabelsi@surys.com

***Abstract*—Onboarding of new customers is a sensitive task for various services, like Banks who have to follow the Know Your Customer (KYC) rules. Mobile Onboarding Applications or KYC by Streaming are expanding rapidly to provide this capacity at home. Unfortunately, this leaves the authentication tools in the hand of end-users, allowing the attacker to directly tamper the video stream. With the rise of new digital face manipulation technologies, traditional face spoofing attacks such as presentation attacks or replay attacks should not be the only one to be considered. A new kind of face spoofing attacks (i.e. digital face spoofing) needs to be studied carefully. In this paper, we analyze those new kinds of attacks and propose a method to secure identity documents against both the traditional attacks and the new ones.**

***Index Terms*—Identity theft, remote onboarding, KYC, image forgery, cryptographic seal, image authentication, portrait seal, perceptual hash, Face anti-spoofing.**

I. Introduction

Remote Biometric Onboarding (RBO) includes all the methods allowing to collect and secure data concerning a not-previously-known applicant. The applicant uses his/her own smartphone or computer to communicate with the service.

Secure RFID chips are only available on few ID documents (mainly passports) and the majority of smartphones or computers are not able to communicate through RFID. Remote identification thus relies mostly on images or video processing.

Images of the ID documents can either be acquired by a webcam or a smartphone. The same device is then used to acquire a video (or a series of photo) of the document holder to check if he is the rightful owner of the document.

The remote onboarding process can thus be divided in two steps:

1) Security features verification on the image of the ID Documents: Machine Readable Zone (MRZ) of passports are checked by controlling the consistency with the visual inspection zone (VIZ), patterns of the Diffractive Optical Variable Image Devices (DOVID) are checked and some other minor controls are done.
2) Biometrics Similarity checks: biometrics comparisons are operated between the portrait of the ID Document and the person in front of the camera. Liveness detection process can be added to avoid various face spoofing attacks (e.g. presentation attacks, replay attacks, etc).

Until recently most of the attacks against this process came from physical counterfeited or fraudulent ID documents, but a variety of new technologies providing photorealistic computer-generated images bring new ways of fooling RBO. New processes and tools must be defined to provide secured RBO against those emerging threats.

II. Categories of attacks

A. Physical attacks and counter-measures

To deceive the RBO process, fraudsters have several physical possibilities to spoof the system:

- Present a face printed on paper instead of his face: liveness detection has been conceived to detect such kind of presentation attacks
- Use a mask to mimic someone else, with the possibility to close the eyes or open the mouth to succeed in liveness tests: color reflection on the skin/mask or lack of facial activity can be analyzed to detect it.
- Use a high-resolution screen to display an animated version of the target face: interactive sessions or moiré-artifacts analysis [1] will unveil the fraudster.
- Buy a fake ID document with a look-a-like person or obtain a fraudulent one with a fake photo [2].

All physical attacks, except picture replacement, require:

- Expensive components (mask, Fake ID)
- Risks (to buy a fake or mask)
- Delays (to organize their use)

In comparison, digital attacks are a lot easier but maybe a bit riskier for the fraudster. Indeed, in the case of a face morphing attack, the attacker will inject part of his identity into the image.

B. Digital attacks

Digital attacks take profit of one weakness of the biometric system. As seen in Fig. 1, the image or video comes from a physical sensor. At this point, a digital representation of a real-world scene is generated and send to the software which extracts features and performs biometrics matching. Nowadays, images and videos coming from the sensor are not watermarked or signed in any secure way and thus cannot be authenticated.

In a case where the application is hacked, an attacker could control the camera stream and inject a forged image or video

This work was supported by the ANR project DEFACTO ANR-16-DEFA-0002 and by the French General Delegation for Armaments (DGA).

EISIC 2019, November 26-27, 2019, Oulu, Finland
978-1-7281-6735-0/19/$31.00 © 2019 IEEE

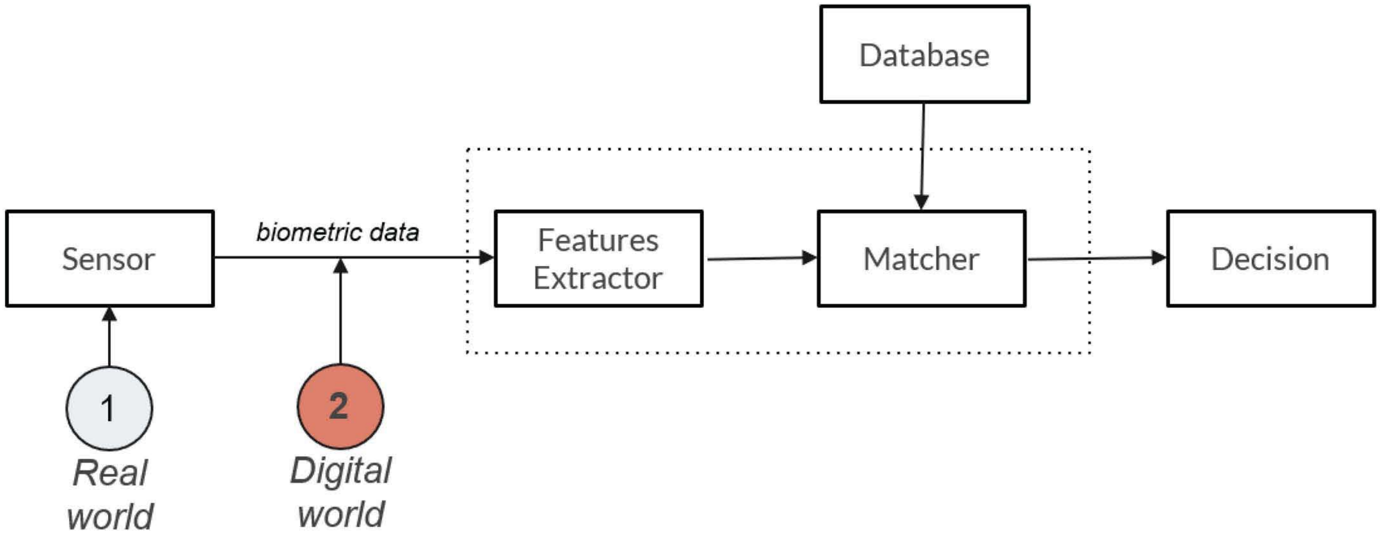

Fig. 1: The two main attacks on RBO

stream before the software process. With current architectures, and at least as long as the camera is not part of the secure elements of the smartphone, this type of attack cannot be prevented.

Once the camera stream is under the control of the fraudster, several possibilities are opened:

1) Many technologies allow to transform the characteristics of a 3D face on a video stream to look like someone else. Deepfake [3] is probably the best-known method to do this process and today it is becoming possible to create a deepfake without a subject specific training [4], Portrait Animation [5] allow the reenactment of a single portrait from a live face acquisition, Face2Face [6] is one of the most efficient allowing real-time animation of a target face from a facial live capture. It allows reenacting in real-time to liveness challenges required by the service. If the user is asked to smile, the fraudster will smile and Face2Face algorithm will reenact the mouth of the target. In [7] extends the quality of such methods by fully re-rendering the person face based on an initial texture. This allows to produce fine scale details of the skin texture.
2) A second set of technologies focus on the digital real-time replacement of the 2D portrait picture on ID Documents. Face replacement in ID documents is easy to operate in real-time, it is sufficient to detect the precise position of the portrait face and the pose vector of the ID Document to be able to inject the replacement face in an acceptable way. It is necessary after this step to correct the local balance of colors of the injected face in order to keep the coherency with the original document, in particular to keep realistic the holographic security elements which are covering a large part of the picture on most ID Documents. We have developed tools to operate such replacement in real-time Fig. (2) and have also constructed a reference database of doctored images based on various face replacement strategies (morphing, swapping, inverse fit swapping).

The first set of technologies are very promising for counterfeiters because they do not compromise the face of the fraudster, but in practice those methods are not easy to manipulate, and it is still difficult to maintain a real-time production with high-resolution video.

C. Analysis of counter-measures

Methods of fake detection based on neural networks have been proved efficient [8], in particular with the MesoNet approach [9] or the FaceForensics++ approach [10], but it is not yet clear how those methods are robust to the evolution of the deepfake generation process. So this area does not appear as sufficiently mature.

The second set is theoretically available to potentially any persons with basic informatics knowledge, and this will be our main concern in this paper.

As mentioned in the previous paragraph, passive forensic methods, which try to detect the image or video stream forgeries without any indication may not be sufficient against an always evolving threat. We believe that active forensic methods, which secure the media by watermarking it or digitally signing it, are more adapted and more robust against face and ID spoofing attacks.

D. Portrait seal

How to prove the authenticity of a picture? More specifically in our case of a portrait? In a fully digital scenario, there are many ways to protect an image. A wide range of watermarking methods can be employed to secure a file. Fragile watermarks can ensure that no information in the image content has been altered and can locate any alteration performed. Another common method to authenticate a file is to digitally sign it. A

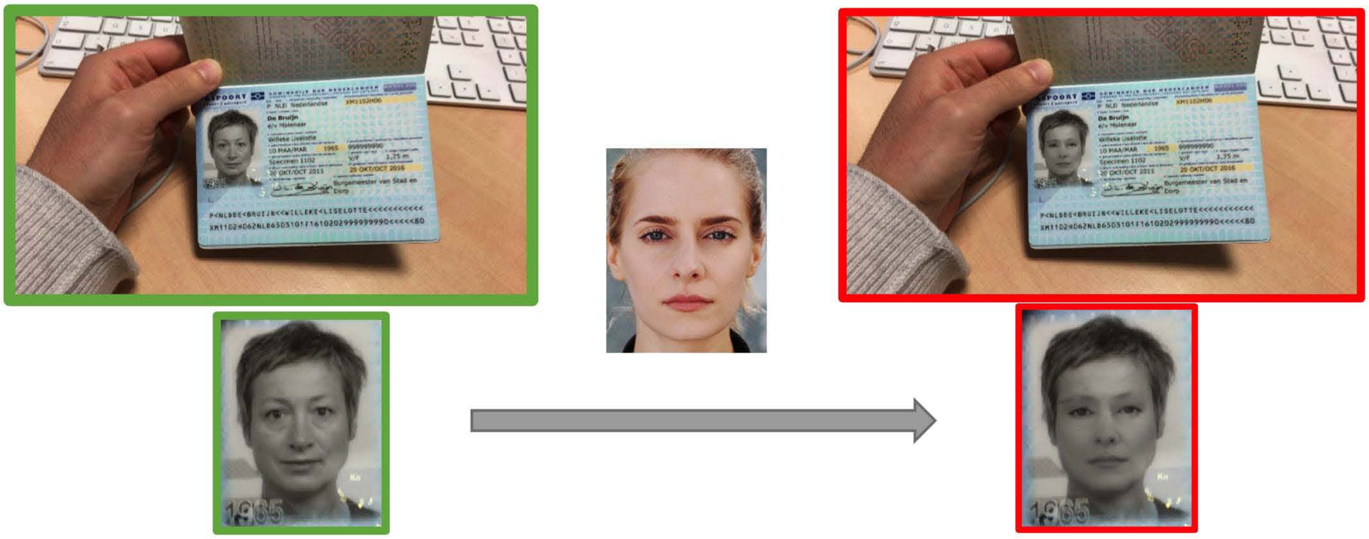

Fig. 2: Example of Real-Time video portrait inverse fit swapping.

signature will secure the complete file (including meta-data) but will not allow to locate modification on the file. All those methods are commonly used as they are known to be extremely robust.

In our case the picture is a printed portrait on physical ID document. In that case, no digital signature algorithms can be applied and watermarking techniques become less efficient.

We have proposed in [11] a method to authenticate a portrait after a Print-Scan or Print-Cam process. At enrolment stage, a set of salient point is extracted from the portrait. The location of those points are stored in a 2D visible seal and digitally signed using a public-key cryptography strategy [12], [13]. The authentication of the portrait is then performed by comparing the stored locations against the recomputed locations on the reacquired picture.

This method ensures that the biometric traits of the portrait have not been altered in any way while not being a biometric fingerprint. In fact, two portraits of one person will not match using the method in [11].

One drawback of this method is that it does not authenticate the content of the portrait. It only ensures that the structure of the portrait is similar to the one stored in the seal. One attack against the system is thus to deform the structure of a face portrait B so that it matches the structure of the face portrait A stored in the visible seal.

To further secure the authentication process, we proposed an extension [14] of [11] to also authenticate part of the content of the portrait. A perceptual hash of the pixels intensity on a path between the salient points is computed and stored in the visible seal. The path between the salient points can be of any type, but should preferably cover important biometric traits of the portrait (eyes, nose ...).

This added constraint in the authentication process make the previously described attack even more difficult to achieve. To fool the system, the attacker should create a portrait that both match the structure and the content of the original portrait. While generating a portrait that fool the system does seem trivial, one has to remember that the generated portrait should also look like the attacker which in practice makes the attack intuitively hard as in [15].

As for [11], the extension [14] is still not a biometric fingerprint of the person as two distinct portraits of the same person would still not match. This is an important property of the method as it is GDPR [16] compliant.

A potential implementation of the seal requires fewer than 100 bytes with usual ECDSA cryptography, which can be easily embedded in a QR-Code or a Datamatrix added to the physical document or store in a separate database or file.

III. Experimentation corpus

A. Attacks

Four main types of attacks have been evaluated:

1) Complete Photography Replacement: Is the basic case which requires to be evaluated because it is a very common and easy attack. The original photography is simply replaced by the attacker portrait. In this case the complete biometry is altered.

2) Face Swapping: Is similar to the complete photography replacement but only the facial area is replaced. This has the advantage of being more subtle than a complete replacement and might be harder to detect. In this case also the complete biometry is altered.

3) Face Morphing: Is the fusion of the biometric identities of two (or more) different people in order to create a portrait with shared biometrics. A document holding such a picture could allow both people to use it indifferently. This attack has first been shown to be a threat in [2] and has since been studied with great attention.

TABLE I: Area Under the Curve (AUC) for multiple attacks

Attack	Photo replacement	Face Swapping	Face Morphing	Identity deletion
AUC	99.70	99.71	99.98	99.85

4) Identity Deletion: Consists in altering the biometric characteristics of one person in order to erase the possibility to recognize him/her or to erase some specific identifiable traits of this person. Currently this technique, of growing interest [15], is mainly used for preserving privacy.

B. Creation

Original face images were taken from the FERET dataset [17], [18] and gathered from IMDB. Face morphing and Face swapping were generated as in [19] with $\alpha = 0.5$ for morphing and $\alpha = 1$ for swapping. For identity deletion, the facial image A is warped to match the biometric traits of B as in [19] but no blending is performed afterwards.

After each attack, random post-processing (gamma adjustment, blurring, rotation, translation, shearing, scratches) are applied to simulate the Print-Scan and Print-Cam process. An example of degradation is shown in Fig. 3. Those random post-processing are also applied to the original images.

IV. Results

To assess the performance of our method, experiments were conducted on the dataset described in III. Every image in the dataset is compared against every other images. The Area Under the Curves (AUC) are reported in Table I. It can be seen that our method as very little chances to falsely match two different portraits even under Print-Scan and Print-Cam scenarios (lowest AUC is 99.70%). The difference in performance against different attacks is due to the randomized post-processing described in III-B.

V. Conclusion

Remote Biometric Onboarding will become widespread in the coming years in various categories of service. Identity theft will rely on hacker's tools and photo-realistic image doctoring. To fight such risks, two strategies will be necessary: the detection of digital fakes and the usage of light-weight content seal.

We demonstrate in this paper that portrait seals can be applied to protect the authenticity of the face images present on ID Documents with a level of confidence higher than human capacity to recognize the document holder. Such Portrait Seal can be embedded in the original document as a simple QR-Code or Data-matrix or in a more sophisticated pattern like Photometrix [11].

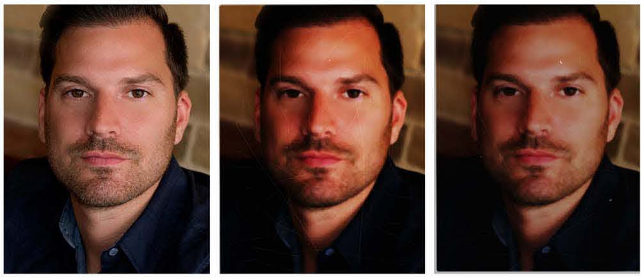

Fig. 3: Random post-processing applied on all images

References

[1] K. Patel, H. Han, and A. K. Jain. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security*, 11(10):2268–2283, Oct 2016.

[2] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. The magic passport. *IJCB 2014 - 2014 IEEE/IAPR International Joint Conference on Biometrics*, pages 1–7, 2014.

[3] What are deepfakes why the future of porn is terrifying. https://www.highsnobiety.com/p/what-are-deepfakes-ai-porn/. Accessed: 2019-10-25.

[4] Yuval Nirkin, Yosi Keller, and Tal Hassner. FSGAN: Subject Agnostic Face Swapping and Reenactment. 2019.

[5] Hadar Averbuch-Elor, Daniel Cohen-Or, Johannes Kopf, and Michael F. Cohen. Bringing portraits to life. *ACM Trans. Graph.*, 36(6):196:1–196:13, November 2017.

[6] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner. Face2Face: Real-time Face Capture and Reenactment of RGB Videos. In *Proc. Computer Vision and Pattern Recognition (CVPR), IEEE*, 2016.

[7] Justus Thies, Michael Zollhöfer, and Matthias Nie. Deferred neural rendering: Image synthesis using neural textures. *ACM Trans. Graph.*, 38(4):66:1–66:12, July 2019.

[8] Belhassen Bayar and Matthew C. Stamm. A deep learning approach to universal image manipulation detection using a new convolutional layer. In *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec '16, pages 5–10, New York, NY, USA, 2016. ACM.

[9] Darius Afchar, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. Mesonet: a compact facial video forgery detection network, 2018.

[10] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. FaceForensics++: Learning to Detect Manipulated Facial Images. 2019.

[11] M. M. Pic and A. Ouddan. Photometrix (tm): A digital seal for offline identity picture authentication. In *2017 European Intelligence and Security Informatics Conference (EISIC)*, pages 163–163, Sep. 2017.

[12] Public key cryptography for the financial services industry: The elliptic curve digital signature algorithm (ecdsa). ANSI X9.62-2005, American National Standards Institute, 2005.

[13] Standards for efficient cryptography, SEC 1: Elliptic curve cryptography. Version 2.0, Certicom Research, May 2009.

[14] M. M. Pic and Gaël Mahfoudi. A phygital vector for identity, robust to morphing. In *Digital Document Security*, Berlin, Germany, 2019.

[15] Debayan Deb, Jianbang Zhang, and Anil K. Jain. AdvFaces: Adversarial Face Synthesis. 2019.

[16] European Commission. *2018 reform of EU data protection rules*. https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf.

[17] P.Jonathon Phillips, Harry Wechsler, Jeffery Huang, and Patrick J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295 – 306, 1998.

[18] P. J. Phillips, Hyeonjoon Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, Oct 2000.

[19] Gaël Mahfoudi, Badr Tajini, Florent Retraint, Frédéric Morain-Nicolier, Jean Luc Dugelay, and Marc Pic. DEFACTO: Image and Face Manipulation Dataset. In *27th European Signal Processing Conference (EUSIPCO 2019)*, A Coruña, Spain, September 2019.