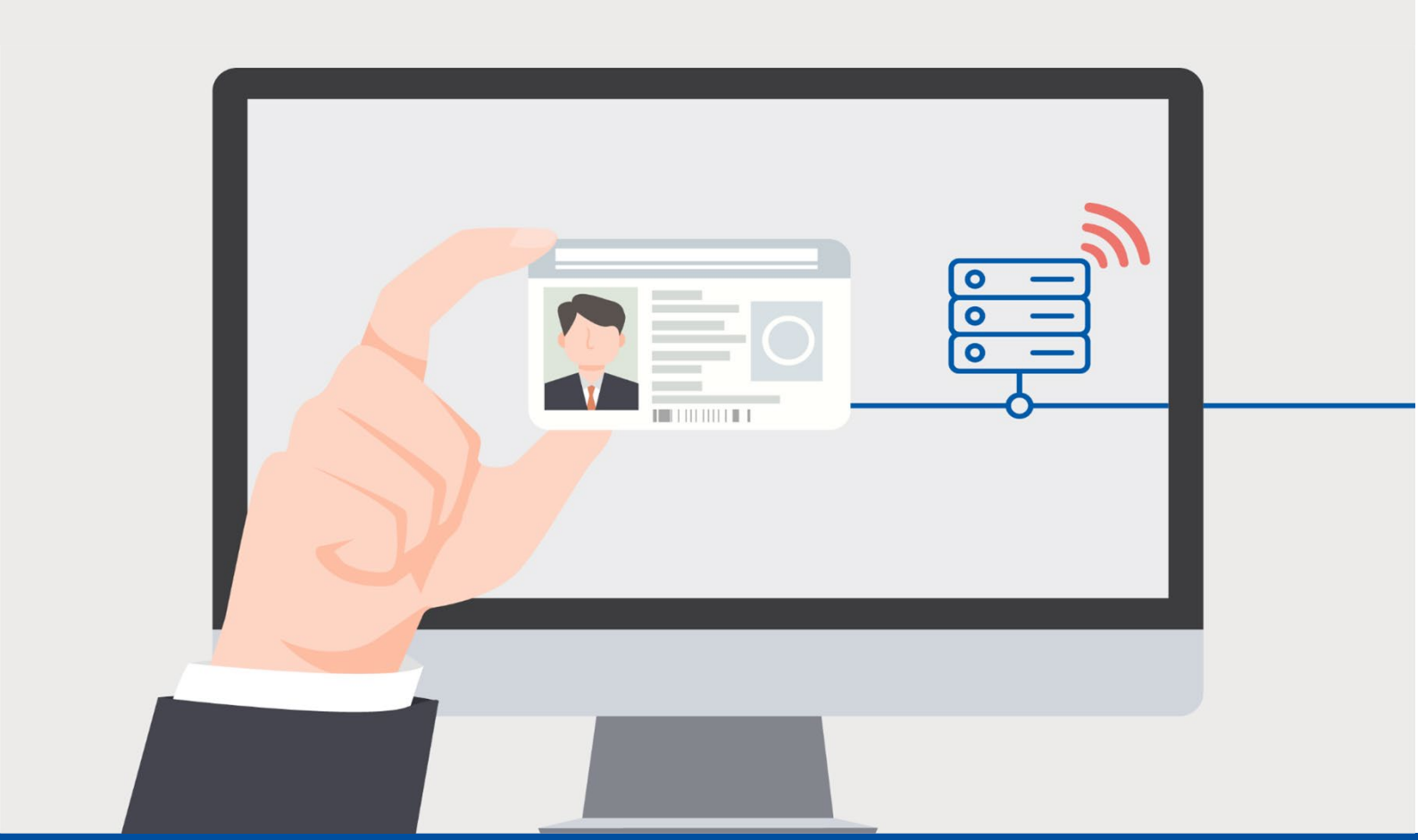
EUROPEAN UNION AGENCY
FOR CYBERSECURITY



REMOTE ID PROOFING GOOD PRACTICES

MARCH 2024

ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services, and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure, and, ultimately, to keep Europe's society and citizens digitally secure. More information about ENISA and its work can be found here: www.enisa.europa.eu.

CONTACT

For contacting the authors, please use eid@enisa.europa.eu
For media enquiries about this paper, please use press@enisa.europa.eu.

CONTRIBUTORS

Athanasios Vrachnos, Evangelia Papadaki, Panagiota Lagou, Nikolaos Soumelidis, Eirini Papamichail

EDITORS

Evgenia Nikolouzou (ENISA), Rossen Naydenov (ENISA)

ACKNOWLEDGEMENTS

Special thanks go to various stakeholders who provided their response to the survey and/or were interviewed for the purpose of this report, particularly: CLR Labs (Kévin Carta), FaceTec (Alberto Lima, Jay Meier, Kevin Alan Tussy), iProov (Andrew Newell, Campbell Cowie), IDNow (Sebastian Elfors, Rayissa Armata), Veridas (Mikel Sánchez), Yoti (Florian Chevoppe-Verdier, Paco Garcia), Infocert (Lorenzo Piatti, Igor Marcolongo, Leone Riello), Inverid (Bob Hulsebosch), NASK (Daniel Wachnik), SK ID Solutions (Kalev Pihl), Zetes (Daniel Ocampo, Bart Symons), ENISA ECATS Expert Group, European Commission eIDAS Cooperation Network and various public authorities such as the Danish Agency for Digitisation – DIGST (Mogens Rom Andersen), French National Cybersecurity Agency – ANSSI (Andrea Röck, Mathieu Jorry, Lucie Ayala, Mickael Lam), Ministry of Economic Affairs and Digital Transformation of Spain (Beatriz Puerta, Vanesa Sánchez Rojo, María José Villacampa), German Federal Office for Information Security – BSI (Thomas Schnattinger), German Federal Network Agency – Bundesnetzagentur (Konstantin Götze), Polish Ministry of Digital Affairs – CYFRA (Marcin Fijałkowski), Estonian Information System Authority – RIA (Erika Adams) and others across the EU who also contributed to this report.

LEGAL NOTICE

Third-party sources are quoted as appropriate. ENISA is not responsible or liable for the content of the external sources including external websites referenced in this publication.

Neither ENISA nor any person acting on its behalf is responsible for the use that might be made of the information contained in this publication.

ENISA maintains its intellectual property rights relating to this publication.

COPYRIGHT NOTICE

ISBN: 978-92-9204-661-3, DOI: 10.2824/885606

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Over the last decade, an accelerating digital transformation is being observed, which has provided numerous benefits to European society and the economy by facilitating trade and the provision of services, creating new opportunities for businesses and increasing productivity and economic gain. Furthermore, the pandemic highlighted the significance of well-regulated and standardised remote identification processes, along with trustworthy digital identities on which public and private sector organisations may rely. These elements are also emphasised in the planned eIDAS revision (eIDAS 2.0), which will provide all EU citizens with safe and transparent access to a new generation of electronic services, including the EU digital identity wallet (EUDIW). These developments are part of the Commission's wider vision for Europe's digital transformation, Europe's Digital Decade ([1]), setting concrete objectives and targets for a secure, safe, sustainable and people-centric digital transformation by 2030.

Digital identity and identity verification are core functions of most services foreseen in the above context. Therefore, the need for secure and reliable identity proofing services, deployable quickly, at scale and in a cost-efficient manner intensifies, since it is a key enabler for electronic transactions in the Single Digital Market, and due to the increasing volume and sophistication of attacks.

Through this report, ENISA attempts to accomplish the following strategic goals, in the domain of trust services and electronic identification:

- to increase stakeholders' awareness;
- to assist in the risk analysis practices in the rapidly changing threat landscape of identity proofing;
- to contribute to the development of stronger countermeasures, enhancing the trustworthiness and reliability of remote identity proofing (RIDP) methods.

The motivating factors to produce this report were:

- the recent developments in the attack landscape, causing concerns about the trustworthiness of identity proofing;
- requests from various stakeholders regarding up-to-date information and guidance on defensive good practices.

Based on the above, the scope of this report builds and expands on the 2022 ENISA report *Remote Identity Proofing – Attacks & Countermeasures* ([2]), in an effort to bring novel types of threats and wider ecosystem concerns to the foreground.

The information and data analysis phase, which consisted of a literature review, two surveys and subsequent rounds of interviews, identified the following major attacks:

- biometric presentation and injection attacks against a human subject's face;
- presentation and injection attacks against an identity document.

1 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en.
2 https://www.enisa.europa.eu/publications/remote-identity-proofing-attacks-countermeasures.

Consideration was given to the nature and developments relating to deepfake attacks and related approaches in offensive and defensive aspects.

Next, applicable countermeasures were analysed and proposed as a set of good practices in the domains of environmental, procedural, organisational and technical controls. The rate and sophistication of novel threats require a revised mindset of defence, incorporating preventive and detective approaches.

The report briefly examines attacks relating to identity documents that take place during the evidence validation and information binding phase of RIDP. The two most prominent good practices for defending identity documents were the status lookups in various identity document registries and the scanning of the near-field communication (NFC) chip (where available).

Both practices have their own obstacles in the course of their full realisation. Many of the identity document registries are maintained on a voluntary basis and a central, up-to-date registry with all the latest document versions of each Member State does not currently exist. On the other hand, while scanning the NFC chip to verify the holder´s personal information and biometric photo could eliminate several of the synthetic attacks, it is not currently legally and consistently permitted for private entities (trust service providers (TSPs), RIDP providers) across the EU. The inconsistent state of NFC-reading can be thought of as a part of the wider scattered regulatory landscape across the EU relating to the recognition of the remote nature of identity proofing and the assurance level it can provide.

Finally, the report highlights wider concerns of the landscape, unrelated to attacks or technical topics, but capable of affecting the secure adoption and execution of RIDP methods across the EU.

1. INTRODUCTION

1.1 CONTEXT

The purpose of this study is to build upon previous ENISA studies on RIDP and focus on new developments, security recommendations and good practices, when RIDP is used in the context of the eIDAS regulation, the 6th EU anti-money laundering directive or any other context where trust in the identity of a natural or legal person is essential.

Identity verification in Europe is undergoing a period of intense transformation. Since the COVID-19 pandemic outbreak, RIDP has been under an intense evolution; from face-to-face verification in stores to synchronous and asynchronous remote identity document and biometric verification that can also be processed automatically. The upcoming eIDAS 2.0 regulation and the introduction of the EUDIW, with the ambition that 80 % of EU citizens will make regular use of the wallet by 2030, will extend the cases requiring identity verification with a high level of assurance. Similar developments are also taking place at the global level, with numerous efforts to design and develop decentralised digital identity wallets.

There is a significant increase in demand for secure, reliable and user-friendly RIDP, for the following reasons.

- Identity proofing is a key enabler for electronic transactions and the development of the digital single market across the EU. This becomes even more significant with the upcoming eIDAS revision and the EUDIW, which will enable access to a new breed of electronic services in a secure and transparent way for all EU citizens.
- The increasing volume and sophistication of attacks causes concerns regarding the trustworthiness of the process. This relates to the emergence of new types of attacks, such as high-quality deepfakes, and the availability of computational resources and tools which allow scaling and automation.

1.2 SCOPE

The goal of this report is to provide an updated, inclusive view on attack techniques against RIDP mechanisms, validate the security controls proposed in the previous ENISA report for presentation attacks and provide further practical countermeasures to mitigate new types of attacks.

The study falls under ENISA's efforts to support:

- implementation of the eIDAS regulation by addressing technological aspects and building blocks for trust services, electronic identities and digital wallets;
- analysis of the cybersecurity requirements stemming from the Commission recommendation to develop a common EU toolbox for a coordinated approach towards a European Digital Identity Framework (eIDAS 2.0);
- the development and implementation of EU policy in the field of electronic identity and trust services per the EU Cybersecurity Act mandate.

The study covers new developments in the area of RIDP, such as digital injection attacks and the various methods to conduct them, and provides additional insight with regards to security requirements and good practices collected by various stakeholders, such as research institutions and academia, the industry (identity proofing software vendors and service providers, TSPs, etc.) and gatekeepers (conformity assessment bodies, supervisory bodies).

EXTENDED SCOPE

This report has an extended scope in comparison with the previous ENISA report on RIDP, covering presentation and injection attacks against the face and identity documents.

Aspects that have not been analysed in the current report:

- human-related, internal threat scenarios regarding operators of an RIDP system, such as a disgruntled employee helping the attacker by tampering with internal data or a deceived employee who has fallen prey to a social engineering attack;
- attacks focusing on biometric elements other than the face (e.g. voice, fingerprints);
- attacks relating to earlier phases of identification (e.g. enrolment of ID documents), such as morphing attacks and related defensive approaches (e.g. morphing attack detection (MAD));
- generic cyberattacks aimed at underlying technologies (user workstations and servers, tampering with data in transit given improper encryption, etc.) or human factors (generic social engineering attacks, etc.), except where a close and direct impact on RIDP methods is specifically observed and explained;
- the tools, means and sources to accomplish the illustrated attack scenarios;
- detailed technical or scientific analysis of offensive tools and methods;
- feasible but implausible scenarios, such as attacks performed by doppelgangers (often seen among family members) or plastic surgery to impersonate someone else, etc.;
- benchmarking/evaluation between in-person identity verification and RIDP methods;
- privacy and data protection issues related to personal data, biometric data processing, etc.

However, the fact that these aspects have been excluded does not imply they should be ignored when performing a risk analysis and wider evaluation, prior to onboarding or developing an RIDP solution.

1.3 METHODOLOGY

The collection and analysis of information was of paramount importance in this study. This required inventorying sources of information and potential participants/contributors based on their roles and responsibilities, expertise on the topics, active roles in developing new technological means to combat spoofing attempts and available publications proving real experience in this relevant field, ensuring critical review of the collected information by the study team and applying proper quality assurance methods.

For these reasons, the project team prepared the necessary tools to support its approach and methodology, i.e. creating databases of information sources and targeting stakeholders, designing and implementing the questionnaire for the survey, collecting feedback on draft deliverables and comments received during the workshop and allowing end-to-end traceability of the information collected and incorporated into this report.

A combination of desk research, interviews, surveys and workshops was used to collect data. The vast amount of information was elicited through an online survey. The results were analysed and aggregated based on qualitative and quantitative methods, thus drafting preliminary insights and ensuring the confidentiality of the information.

Follow-up interviews were conducted with interested stakeholders of different categories (RIDP providers, TSPs, conformity assessment bodies, laboratories, etc). Customised questions were used during the interviews to gather their specialised experience; this allowed a more detailed analysis of new threat scenarios and possible countermeasures as perceived by field experts.

1.4 TARGET AUDIENCE

The present report is aimed primarily at the following stakeholders.

- **Supervisory bodies** that supervise or approve the certification of RIDP solutions or TSPs using those solutions.
- **Conformity assessment bodies** that evaluate RIDP solutions or TSPs using the solutions.
- **TSPs and identity providers** that might use this report to strengthen cybersecurity of their own RIDP solutions.
- **Security researchers, academia, students, laboratories** and the wider security community.
- **Companies and other private sector organisations** that run or are preparing to adopt an RIDP solution for customer onboarding.
- **Governments and various public bodies** that are considering implementing an RIDP solution for citizens, employees and other stakeholders.

1.5 STRUCTURE

The logical structure of this report begins with an updated overview of the existing RIDP methods and the literature review on recent developments in national legislation across the EU and applicable standards (Chapter 2). The report continues with presenting an up-to-date view on RIDP attack methods based on the current state of the threat landscape (Chapter 3), followed by proposed countermeasures (Chapter 4). Finally, conclusions and wider concerns of the landscape, unrelated to attacks or technical topics but capable of affecting the secure adoption and execution of RIDP methods across EU and conclusions, are drafted in the last chapter (Chapter 5).

1.5.1.1 Chapter 2. Background

This section provides an updated view on the recent development that have been made since the publication of the last ENISA report of 2022 on Remote Identity Proofing Attacks & Countermeasures, including certification schemes and legislation published by Member States, along with updated publications and works in progress of European and international standards on biometric security and biometric testing.

1.5.1.2 Chapter 3. Attack overview

Regarding attack methods, the analysis performed in this chapter is based on the latest insights of the observed attack types and whose feasibility has been validated by security researchers, but also includes those types still in a conceptual stage, but with a realistic probability of being introduced in the future. The study focuses on **presentation** and **injection** attacks against the **human face** and **identity documents**. Attacks are analysed following the instrument-method-terminology, trying to provide validity and consistency with well-known standards on this topic.

1.5.1.3 Chapter 4. Good practices

Good practices and countermeasures are presented according to the following categorisation: face presentation and injection attack detection (IAD) controls, identity document controls, procedural and organisational controls.

1.5.1.4 Chapter 5. Conclusions

The report concludes by capturing some wider concerns and potential obstacles to the secure adoption of RIDP methods across the EU, based on the opinions of various stakeholders of the ecosystem in conjunction with the interaction and alignment happening among the eIDAS regulation, national legislation of the Member States and applicable standards.

1.5.1.5 Supplementary material

Annexes A, B and C contain additional material relating to the methodology and the main results of the interviews, surveys and workshops, along with an updated list of national legislations of Member States on RIDP and a summary of all countermeasures proposed in this document.

ABBREVIATIONS

ANSSI	National Agency for Information Systems Security of France
API	Application Programming Interface
CAB	Conformity Assessment Body
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CaaS	Crime-as-a-Service
CEN	European Standardization Committee
CNN	Convolutional Neural Network
EBA	European Banking Authority
eIDAS	Regulation (EU) 910/2014
ETSI	European Telecommunications Standards Institute
EUDIW	European Union Digital Identity Wallet
FAR	False Acceptance Rate
GAN	Generative Adversarial Network
IAD	Injection Attack Detection
IAI	Injection Attack Instrument
IAM	Injection Attack Method
ISO	International Standards Organization
IVR	Interactive Voice Response
MAD	Morphing Attack Detection
MRZ	Machine Readable Zone
NFC	Near-Field Communication
OCR	Optical Character Recognition
PAD	Presentation Attack Detection
PAI	Presentation Attack Instrument
PRADO	Public Register of Authentic Identity and Travel Documents Online
QWAC	Qualified Web Authentication Certificate
RASP	Runtime Application Self-Protection
RIDP	Remote Identity Proofing
ROFIEG	European Commission's Expert Group on Regulatory Obstacles to Financial Innovation

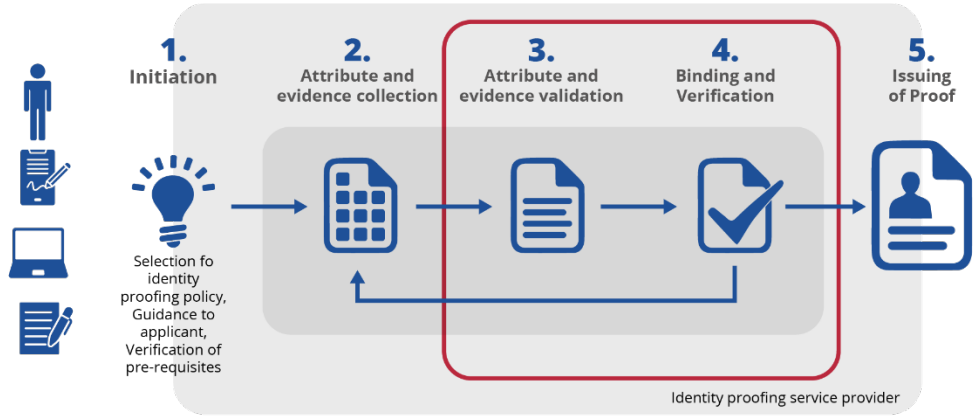
SLTD	INTERPOL's Stolen and Lost Travel Documents database
TSP	Trust Service Provider
TEE	Trusted Execution Environment
SB	Supervisory Body
SDK	Software Development Kit
SOC	Security Operations Centre

2. BACKGROUND

2.1 INTRODUCTION

The generalised five-step RIDP process was initially presented in the 2021 ENISA report on the various RIDP methods and is presented below.

Figure 1: Generalised RIDP process



To go through this process, an applicant requires:

- definitive proof (usually photo or video evidence of their face) that the applicant is physically present in front of the capturing device and is physically holding their identification document, during the whole proofing process;
- an official identification document issued by an authoritative source, typically in the form of a government-issued ID or passport;
- a high confidence match between the liveness-proven photo or video evidence and the face shown on the identity document.

Once the match is confirmed, the RIDP provider can bind the document, including all the data it contains, to the applicant (binding phase). Consequently, an attacker whose goal is to fool the system by impersonating someone else has the choice of spoofing:

- the identity document, for example by forging the photo part of an authentic identity document;
- their face, by forging photo or video evidence to match with the one on an authentic document;
- both, by using a fake document and fake face evidence.

Although this report includes updated information regarding identity spoofing using forged identification documents, the topic of the production or acquisition of such documents and related weaknesses in the process is omitted. The report focuses mainly on attacks that target the binding phase by spoofing the applicant's face.

2.2 SUMMARY OF REMOTE ID PROOFING METHODS

RIDP methods can be categorised as follows. Please note that the methods included below are not considered equivalent in the level of security and assurance they may provide.

Method	Description
Videocall with operator	The applicant submits personal information and is then interviewed by a human operator through a video conferencing system.
Videocall with operator, assisted by software	A videocall is conducted with a human operator, while software (including AI) may be used to assist or streamline the collection of information. The human operator conducts the process and takes the decisions.
Fully automated with photo/video	The applicant's identity is verified through face photo/video with fully automated methods, without any intervention by a human operator.
Fully automated with photo/video, reviewed by operator for low score	The applicant's identity is verified as in the 'fully automated' method; a review by a human operator takes place in cases where the assurance level of the recognition software falls below a pre-established threshold.
Unattended with photo/video, reviewed by operator (hybrid)	The applicant's identity is verified as in the 'fully automated' method; a review by a human operator takes place in all cases.
Electronic identification means	The electronic identity provided by identity providers (e.g. financial institutions) is used, taking advantage of the identity proofing already performed by them in the past.
X.509 certificate-based	This method is based on an internal or third-party trust service, where evidence is based on the possession of the private key of an X.509 certificate, and the person's identification data is retrieved from the certificate, issued through a physical face-to-face process.
Combined methods	A combination of any of the above methods in a single identity proofing process, in order to increase security and the level of assurance.

2.3 REFERENCE TO PREVIOUS ENISA STUDIES & RESULTS

Two reports on RIDP have already been published by ENISA and provided the basis for the current report.

- The **2021 ENISA report** ([3]) provided an overview of the most common methods used for identity proofing; presented the current legal and regulatory landscape and supporting standards at the international and EU levels; discussed the input received though questionnaires from different stakeholders which use, offer or evaluate identity proofing solutions; presented an initial gap analysis on existing standards and regulations; and stressed the need for a harmonised adoption of RIDP and provided a number of legal and technical recommendations.
- The **2022 ENISA report** ([4]) focused on potential threats to RIDP methods, along with the corresponding security controls and countermeasures. Building on the 2021 report, it focused on possible face presentation attacks against RIDP methods. Through the analysis, major face presentation attacks were identified, such as photo attack, video of user replay attack, 3D mask attack and deepfake attack.

2.4 DEVELOPMENTS IN LEGAL & REGULATORY REQUIREMENTS

New regulations come in on a somewhat regular basis, requiring providers and other stakeholders to continuously update their practices for compliance purposes. However, the rate of changes in technology and threat landscape outperforms the legislative cadence. This situation leads to regulations or standards with relatively short lifetime which need to be

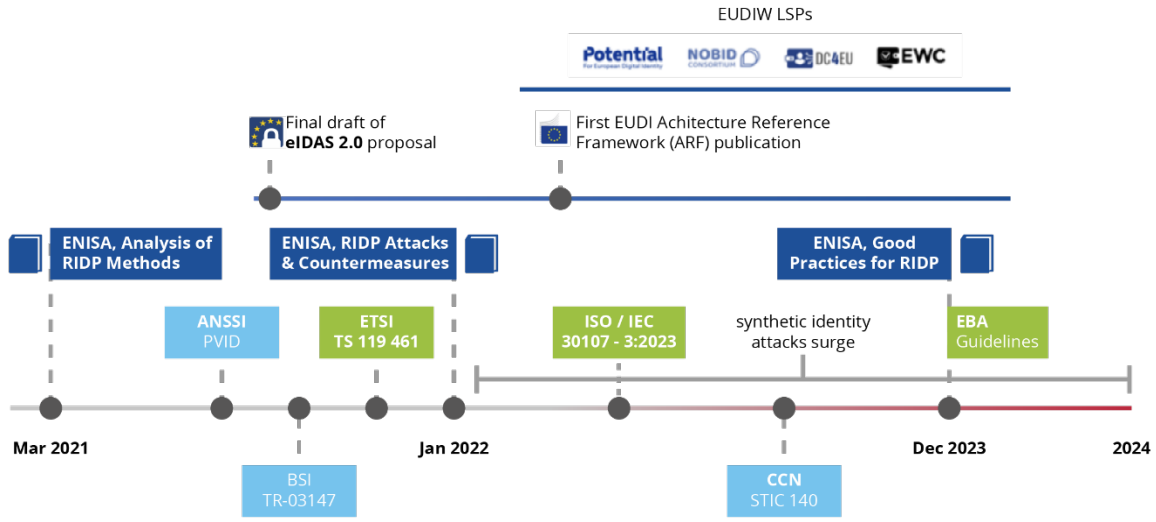
([3]) ENISA, *Remote ID Proofing – Analysis of methods to carry out identity proofing remotely*, March 2021, https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing.
([4]) ENISA, *Remote Identity Proofing – Attacks & countermeasures, January 2022,* https://www.enisa.europa.eu/publications/remote-identity-proofing-attacks-countermeasures.

regularly updated to keep up with the current technology and threat evolution. Moreover, extra pressure is put on service providers to comply, as user expectations evolve and new security concerns arise.

Another key concern is the discrepancy in national regulations among Member States. Despite being under the umbrella of EU regulations, Member States still have different certification schemes for TSPs, leading to a fragmented regulatory and standardisation landscape, making them search for alternative ways and become certified wherever the requirements are more favourable. This was one of the key insights elicited through ENISA's workshop in Amsterdam and the subsequent surveys and face-to-face interviews during the information-gathering phase of this report.

This chapter aims to capture important or recent developments in regulatory requirements among Member States, as well as in standardisation efforts, dated within the last 2 years. This particular period (2021–2023) was selected because there have been extensive developments, both in offensive and defensive technologies and in national legislation that affect identity proofing in various ways.

Figure 2: Related events and developments in the last 2 years



Developments in National Requirements

While the rapid evolution of digital applications in both the public and private sectors has grown, the need for regulatory harmonisation within this changing landscape is essential across the EU and the Member States. Although the existing eIDAS regulation provides pioneering rules on electronic identification and trust services, technical and regulatory gaps result in fragmented provisions across Europe. At the EU level, two new draft proposals have been released:

- a draft proposal for EU regulation No 910/2014 on electronic identification and trust services (eIDAS); and
- a draft proposal for a new European AML regulation.

2.4.1.1 FRANCE

ANSSI, the National Information Systems Security Agency of France, published in March 2021 the PVID standard ([5]) on requirements and recommendations for remote identity verification service providers, verifying the identity of natural persons. These requirements constitute the basis of a certification scheme for identity proofing services, regardless of the level of assurance (whether it is substantial or high), of the proofing method (asynchronous, synchronous, with or without human interaction) and regardless of the applicable regulatory framework.

Important requirements for service providers are:

- annual risk assessment and treatment, also considering identity theft risks;
- definition of a remote identity verification policy and practice;
- data protection;
- organisation and administration of the service provider;
- quality and service levels.

Especially for the risk assessment part, service providers must identify risk scenarios relating to:

- the forgery of identity documents, either by physical or digital means;
- the alteration of the appearance of the user by digital means, such as:
  - utilisation of 'virtual' digital methods to craft a fraudulent face based on photos or videos,
  - identity spoofing via injection of fraudulent photos or videos of an existing person to replace the data captured during the acquisition phase;
- user face similarity fraud (look-alikes, twins, etc.);
- the influence on the behaviour of the user.

Another requirement is a demonstration of testing performed by competent entities, showing the resistance to identified attacks and testing of the risk treatment plan, in order to ensure the effectiveness of the service to detect attempts to spoof the authenticity of the identity document and the detection of living humans.

Finally, additional provisions are related to the definitions of the required skillset of the service provider's staff and the set of identity documents that a compliant identity proofing system can accept.

ANSSI certified the first remote identity verification providers for natural persons in April 2023.

2.4.1.2 GERMANY

In December 2021, **BSI**, the Federal Office for Information Security of Germany, updated the TR-03147 ([6]) – Assurance Level Assessment of Procedures for Verifying the Identity of Natural Persons – making it possible to evaluate different procedures for (initial) identity verification regarding their trust level and thus to make them comparable. The criteria for the trust level evaluation take into account both the scope and the quality of the measures, along with the threats and requirements for procedures for identity proofing and identity verification of natural persons based on identity documents (e.g. identity card or passport). At the national level, Germany's Trust Services Act (Vertrauensdienstegesetz – VDG) foresees two types of identification:

- other identification methods (e.g. video identification);
- innovative identification methods (e.g. video identification with automated procedure).

([5]) https://www.ssi.gouv.fr/uploads/2021/08/anssi-requirements_rule_set-pvid-v1.1.pdf.
([6]) https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/Technische-Richtlinien/TR-nach-Thema-sortiert/tr03147/tr03147_node.html.

The innovative identification methods are not yet officially recognised and may be provisionally recognised by the Federal Network Agency for up to 2 years, provided a conformity assessment body has confirmed that the identification method has equivalent security as defined in Article 24(1)(d) of Regulation (EU) No 910/2014. Both categories of video identification methods have some limitations by law, as they are considered to provide lower assurance, and thus cannot be used for issuing qualified web authentication certificates; they are only intended for issuing one-time qualified signatures and seals.

2.4.1.3 SPAIN

Regarding electronic trust services, in December 2020 Spain passed the Law 6/2020 ([7]). Per this law, specific mentions are made of 'other conditions and technical requirements for remote identity verification' and 'other identification methods such as video conferencing or video-identification that provide equivalent security in terms of reliability to physical presence as assessed by a conformity assessment body'.

Subsequently, order ETD/465/2021 ([8]) was set into force to regulate remote video identification methods for issuing qualified electronic certificates. The order foresees not only the obligation to use a qualified product of video identification of the CCN catalogue (see paragraph below), but also a set of countermeasures to increase the security level. Highlights of these countermeasures are the following:

- unattended RIDP methods are not permitted (video and evidence are always reviewed by a human operator);
- providers should apply organisational and procedural measures proportional to the risks and appropriate to the nature of the services provided;
- the RIDP system used in the process must incorporate the necessary technical and organisational means to verify the authenticity, validity and integrity of the identification documents used, and to verify the correspondence of the document holder with the applicant performing the process, using technologies such as facial recognition, and to verify that the applicant is a living person who is not being impersonated;
- satisfaction of the above requirements by an RIDP service must be accredited, per Annex F.11 of the CCN-STIC-140 (STIC Security Guidelines), by means of product certification;
- the provider's staff in charge of verifying the identity of the applicant are required to verify the accuracy of the applicant's data, using the captures of the identity document used in the process, in addition to any other automatic means that may be implemented in the remote video identification systems;
- the provider must enforce physical security measures and assess the security of all RIDP system elements (communication channels, creation and storage of evidence, etc.) and training of the staff;
- fulfilment of all these requirements by the provider must be assessed and confirmed by a CAB.

At the technical level, the **National Cryptologic Center of Spain** (Centro Criptológico Nacional – CCN) has developed the STIC Security Guidelines, a series of guidelines and recommendations aiming to enhance cybersecurity within organisations, especially for public administration and companies and organisations of strategic interest. Among these documents, **STIC Security Guidelines CCN-STIC 140 – Annex F.11 – Video Identification Tools,** (published December 2020, last updated March 2022) ([9]), describes the fundamental security requirements of a product from the video identification tool family, in order to be included in the

([7]) https://www.boe.es/buscar/act.php?id=BOE-A-2020-14046.
([8]) https://www.boe.es/buscar/doc.php?id=BOE-A-2021-7966.
([9]) https://www.ccn-cert.cni.es/en/ultimas-guias/5461-guia-140-anexo-f-11-herramientas-de-videoidentificacion/file.html.

qualified products section of the security products catalogue of Information and Communication Technologies, which is also published by CCN. Important topics covered by this document are:

- certifications and assessments required for the qualification of products;
- fundamental safety requirements:
  - protection against replay attacks,
  - biometric verification,
  - audit,
  - secure communications,
  - reliable management,
  - identification and authentication,
  - protection of credentials and sensitive data,
  - validation of the documents submitted.

TSPs are obliged to validate their solutions according to the requirements set out in Annex F.11 of STIC 140.

Developments in Applicable Standards

2.4.1.4 ISO/IEC 30107 multipart standard

The ISO/IEC 30107 series is the most recognised international standard on biometric presentation attack detection (PAD), consisting of four parts:

- Part 1 (2023): Framework;
- Part 2 (2017): Data formats;
- Part 3 (2023): Testing and reporting;
- Part 4 (2020): Profile for testing of mobile devices.

The update of Part 1 in 2023 provides a foundation for PAD by defining terms and establishing a framework through which presentation attack events can be specified and detected so that they can be categorised, detailed and communicated for subsequent biometric system decision-making and performance assessment activities.

The update of Part 3, also in 2023, establishes:

- the principles and methods for the performance assessment of PAD mechanisms;
- the reporting of testing results from evaluations of PAD mechanisms;
- a classification of known attack types.

2.4.1.5 ISO/IEC 29794-5

ISO/IEC JTC1 SC37 WG3 is in the process of revising Part 5 of **ISO/IEC 29794 – Biometric Sample Quality** ([10]), the standard relating to face image data. This part is being developed as an international standard and will be based on quality requirements laid out in ISO/IEC 19794-5:2011 and ISO/IEC 39794-5:2019. The updated standard will cover requirements for software that inspects a single captured image (and not a comparison of multiple images) and definitions of pre-processing methods, along with various quality measures and capture-related quality components which will provide accurate definitions of quality metrics incorporated by quality assessment algorithms.

2.4.1.6 ETSI TS 119 461

([10])https://www.iso.org/home.isoDocumentsDownload.do?t=6jKO2KiKLRpVV9JugAuglHywPUQoX2DyMnOgPthWkOTC9i8BDG23MdV6nIJKTHS3&CSRFTOKEN=SIEZ-DVXZ-DRWG-HV4S-2QQI-3PZF-AQGO-AZG7.

The European technical specification used for the certification of identity providers is **ETSI TS 119 461** ([11]) (Policy and security requirements for trust service components providing identity proofing of trust service subjects), published in July 2021 by the European Telecommunications Standards Institute (ETSI) and defining security best practices for various identity proofing methods, including video, automated, hybrid and NFC-based solutions.

The specification reviews the policy and security requirements for identity proofing and trust services by introducing a new minimum level of identity proofing, which is aligned with the requirements of eIDAS. The standard raises the security level across the identity proofing industry and provides the foundation for issuing qualified certificates and other trust services in the future. Furthermore, ETSI TS 119 461 is being revised and will be updated in accordance with the new eIDAS 2 regulation, which is expected to influence and be a foundation to further regulations, such as the upcoming update to the EU anti-money laundering directive.

2.4.1.7 ETSI GR SAI 011

In June 2023, ETSI also published a group report ([12]) on securing AI usage against manipulation of multimedia identity representations. The report focuses on AI-based techniques of automatic manipulation or creation of synthetic identification data in various media formats. The report also contains highlights of technical approaches and technical and organisational measures to defend against these threats.

2.4.1.8 CEN TC 224/WG 18 – Biometric data IAD

CEN's technical specification is focused on biometric data injection attacks. It is the first attempt to produce a standard on this topic, since a national or international standard for biometric data injection attacks does not currently exist. It aims to provide the basis for IAD by defining terms and establishing a framework through which biometric data injection attacks can be specified and detected, so that they can be categorised, detailed and communicated for subsequent biometric system decision-making and performance assessment activities.

This technical specification provides an overview of:

- definitions of biometric data injection attacks;
- biometric data injection attack use case on main biometric system hardware for enrolment and verification;
- injection attack instruments on systems using one or several biometric modalities.

Guidance is also provided on the following topics:

- a system for the detection of injection attack instruments and injection attack methods;
- appropriate mitigation risk of injection attack instruments and injection attack methods;
- creation of a test plan for the evaluation of IAD systems.

2.4.1.9 EBA guidelines

The guidelines ([13]) on the use of remote customer onboarding solutions set out a common understanding by competent authorities of the steps financial sector operators should take to ensure safe and effective remote customer onboarding practices, in line with applicable anti-money laundering and countering the financing of terrorism (AML/CFT) legislation and the EU's data protection framework.

([11]) https://www.etsi.org/deliver/etsi_ts/119400_119499/119461/01.01.01_60/ts_119461v010101p.pdf.
([12]) https://www.etsi.org/deliver/etsi_gr/SAI/001_099/011/01.01.01_60/gr_SAI011v010101p.pdf.
([13]) https://www.eba.europa.eu/regulation-and-policy/anti-money-laundering-and-countering-financing-terrorism/guidelines-use-remote-customer-onboarding-solutions.

The European Banking Authority (EBA) guidelines on remote customer onboarding will become effective 6 months after their publication in the *Official Journal of the European Union*, i.e. within 2023.

While the guidelines are in principle technology-neutral, in practice they provide directions regarding the technologies that can be applied and those that cannot.

Some of the key points of the guidelines are the following:

- it is possible to use notified eID schemes with a substantial or high eIDAS level of assurance;
- when using identity documents, it must be ensured that the captured photograph or video allows proper verification of the customer's identity;
- liveness detection verifications to ensure the user is present in the communication session;
- foreseen use cases of manual, automated and hybrid identity proofing in an attended or unattended fashion (with or without operator involvement);
- for automated, attended or unattended use cases, liveness detection is recommended, along with other measures against presentation and injection attacks, including deepfakes;
- strong and reliable algorithms to verify that the photograph or video matches the picture originating from the customer's identity document;
- monitor the ongoing adequacy and reliability of the remote customer onboarding solutions (i.e. quality assurance testing, regular automated quality reports, sample testing, manual reviews);
- participation of an employee that has sufficient knowledge of the applicable AML legislation and security aspects of remote verification and who is sufficiently trained to detect and prevent the intentional or deliberate use of deception techniques relating to remote verification;
- develop an interview guide defining the steps of the remote verification process and the actions required from the employee, including guidance on observing and identifying psychological factors, or other features that might illustrate suspicious behaviour.

3. ATTACKS OVERVIEW

This section attempts to present an updated overview of attacks against the RIDP process, based on insights identified during the information gathering and analysis phase conducted prior to publishing this report. This section builds upon the information contained in the 2022 ENISA report *Remote ID Proofing: Attacks & Countermeasures* (ENISA2022), and covers new types of attacks (e.g. data injection attacks) and current evolutions in the threat landscape.

In fact, **deepfake presentation and injection attacks were the top two biometric attack types considered hardest to mitigate** ([14]) by the various stakeholders surveyed during the preparation of this report. Combined with the facts that a significant surge in digital injection attacks has been observed since mid-2022, and that ENISA2022 focused mainly on face presentation attacks, this report was drafted with the objective of providing an up-to-date view on current and evolving attacks; thus, it analyses both presentation and injection attacks, along with an overview of attacks against identity documents.

Recent technological developments in digital image synthesis set the stage for potentially more effective deepfake attack paths. Disentanglement ([15]) is one example of the active research topics in this field, which decouples the facial identity generation process from the pose controlling process, both performed by a generative adversarial network (GAN), allowing to generate results that are customisable and fully controllable, photorealistic to a high degree of quality and natural in facial movement.

GENERATIVE ADVERSARIAL NETWORK
An unsupervised, deep learning model that aims to automatically discover and learn the regularities/patterns in input data, so that the model can generate new examples that can be considered as reasonably originating from the input dataset.

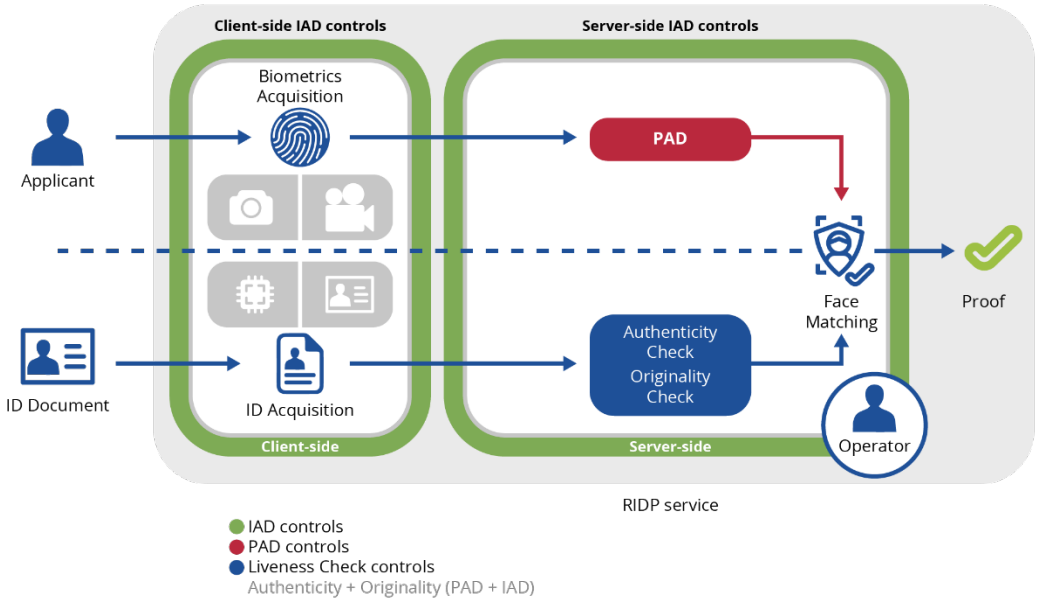
The two core types of technical controls analysed in this report are PAD and IAD controls.

Controls	Description
Injection attack detection (IAD)	A combination of software and/or hardware methods that allow a biometric system to detect spoofing attempts right after the biometrics capture.
Presentation attack detection (PAD)	A combination of software and/or hardware methods that allow a biometric system to detect spoofing attempts during the biometrics capture phase.

To reflect this evolution, a revision of the general RIDP process model has been made, proposing a multi-layer approach in the application of mitigating controls, spanning from the client side to the server side.

([14]) https://www.enisa.europa.eu/events/remoteidentity_workshop_amsterdam2023/remote-id-workshop-amsterdam-briefing.pdf.
([15]) https://www.unite.ai/disentanglement-is-the-next-deepfake-revolution.

Figure 3: Revised general RIDP process model



The revised model contains the following changes, compared to Figure 3 of ENISA2022:

- PAD and IAD controls are in place for the biometrics acquisition (face), along with authenticity and originality checks for identity documents;
- IAD controls cover both the client side and the server side, while PAD controls are mainly considered to be implemented at the server side, at the biometric processing phase, during the biometric data capture of a biometric system.

It should be noted that:

- facial biometric acquisition can be performed either via a photo or video;
- identification information acquisition can be either electronic (e.g. NFC chip scanning) or manual, through optical inspection, depending on the identification document type and its security features.

3.1 PRESENTATION ATTACKS

3.1.1 Overview

Per ISO/IEC 30107-1:2016, **a biometric presentation attack is the specific type that aims to deceive biometric recognition during the biometric data capture of a biometric system.** The face, fingerprints, iris and voice are considered biometric elements. When the attack focuses on the biometric data of the face, it is called a face presentation attack.

Presentation attacks may be performed using one of the following techniques [16].

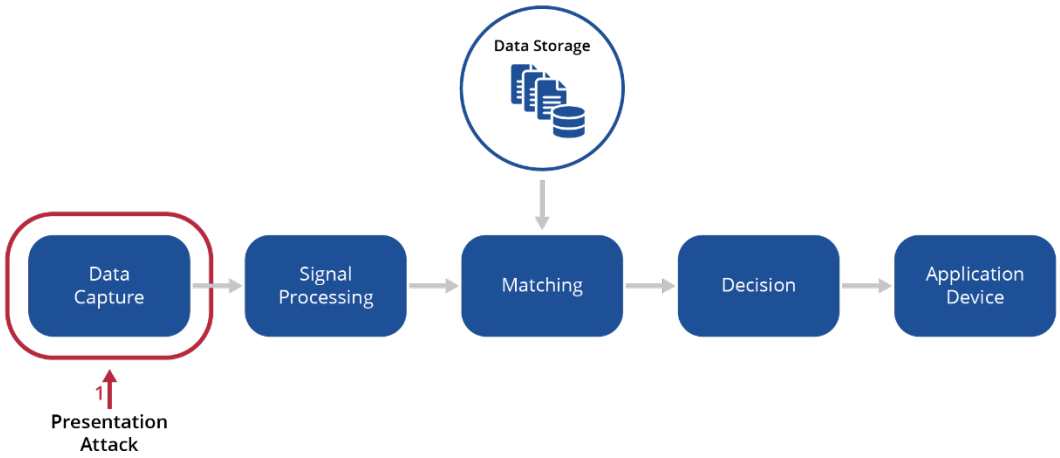
- **Impersonation**. The attacker attempts to copy or mimic physical characteristics and use an identity other than their own, aiming to be recognised:
  - as another **specific individual known to the system** (targeted impersonation), using either direct biometric data from a legitimate user or by fake presentation artefacts, or
  - as **any individual known to the system** (untargeted impersonation).

- **Obfuscation.** The attacker attempts to avoid being recognised by the RIDP system by concealing their own biometric characteristics, through disguise or alteration of natural biometric characteristics (e.g. by using 3D latex masks or facial latex props, extreme makeup) but not necessarily by impersonating a legitimate user's identity. This presentation attack technique is considered less studied [18] compared to the impersonation.

In both impersonation types, recognition is made against an identity that already exists in the RIDP system, whereas obfuscation does not aim to achieve a match against an existing identity in the system.

The following diagram depicts the various processes performed in a biometric system.

Figure 4: The point where presentation attacks take place



Presentation attacks are still an active and evolving threat, even though injection attacks are observed with greater frequency and sophistication. They are considered the two main types of attacks aiming to deceive or modify the biometric sample of a biometrics capture system. The ease and, in turn, popularity of presentation attacks is mainly due to the following factors.

- **Lack of physical presence**. Due to the nature of the RIDP service, lack of physical presence makes it harder for various checks/controls to be adequately enforced. Visibility and observation are limited, and these conditions can easily be exploited by adversaries.
- **Wide range of users.** RIDP services are usually available worldwide, allowing adversaries with varying levels of expertise and motivation to target and exploit them.
- **Ease of conduct.** Some basic types of presentation attacks can be conducted relatively easy by any user, without requiring technical expertise or a deep knowledge of biometrics.
- **Low cost.** Some basic types of presentation attacks do not require expensive or sophisticated equipment, which makes them more appealing for adversaries.

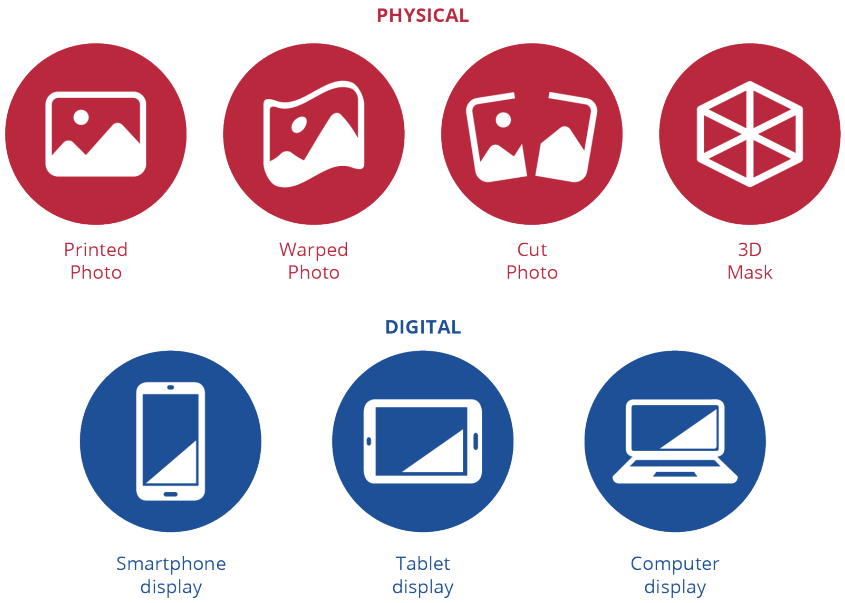
A presentation attack can be considered to consist of two components: the attack method and the attack instrument. While there is a variety of presentation attack instruments (PAIs), the attack method is considered to be only one: the actual presentation of the fraudulent artefact against a camera, utilising an attack instrument. For this reason, the PAI is the only component that is analysed in this report, in an effort to align with the international bibliography on this topic.

Presentation attacks were the main attack topic analysed in the 2022 ENISA report and readers are encouraged to consult this work for a more detailed view on this topic.

3.1.2 Attack Instruments

PAIs are tools utilised to subvert a biometric system by introducing a fraudulent biometric representation of a user during the biometric capture phase (see Figure 5). PAIs can either be physical, such as 2D photographs and plastic and silicone masks, or digital, such as TV, computer tablet and mobile phone screens.

Figure 5: Physical and digital PAIs



Common impersonation attacks utilising the above PAIs are briefly described below. For examples/demonstrations of the various attack instruments, please consult Annex B. Also Annex C provides two real examples from attempted presentation attacks, provided by a QTSP.

3.1.1.1 Printed/Warped Photo

A photo of a legitimate user is printed on paper and placed in front of the face of the attacker, either in a regular or wrapped position (see Annex B, Figure 21), following the attacker's face curvature, so that the camera performing the RIDP process captures the printed face. It is the simplest form of presentation attack, with a relatively low success rate.

3.1.1.2 Printed/Warped 2D Mask

This is a variation of the printed/warped photo attack, where a legitimate face is printed and cut in the form of a 2D mask (see Annex B, Figure 22), paying attention to allow the expression of liveness characteristics by cutting out the eyes and the mouth. It can also be wrapped to follow the attacker's face curvature for slightly more realistic results.

3.1.1.3 3D Mask

Printed 3D layered mask. A photograph of a face is printed multiple times, cut out in the form of a face mask and then all copies are overlayed together, so that it creates a sense of extrusion and face depth (see Annex B, Figure 23).

Hard resin 3D mask. A high-quality 3D mask from hard resin is crafted to mimic the real traits of the human face, also incorporating eye holes, which provide a sense of liveness through eye gaze, blinking and motion. Although hard resin 3D masks achieve spoofing results that are hard to detect by liveness check systems, the material they are crafted from is susceptible to natural limitations, which can be exploited by liveness check algorithms, so that the spoofing attempt is detected effectively (see Annex B, Figure 24).

3.1.1.4 2D Photo to 3D Avatar

With this technique, a 2D face image is used as a reference and fed into a GAN to generate a 3D face puppet, which can then have its facial expressions and characteristics further manipulated. The 3D puppet can then be presented through a screen to the camera performing the RIDP process (see Annex B, Figure 25).

3.1.1.5 Plastic/Latex/Silicone Masks

In this type of presentation attack, a physical mask made from plastic, latex or silicone is crafted and placed on the face of the attacker. Each material provides different qualities and, in turn, different spoofing capabilities, but also has its own level of manufacturing complexity and usually requires additional refinements on the surface of the mask for more realistic results and higher chances for impersonation success (see Annex B, Figure 26).

3.1.1.6 Replay

Photo/video replay. A photo or video of a legitimate user is presented through a digital screen (TV, computer, tablet or mobile phone) to the camera performing the RIDP process. The higher the resolution and pixel density, the higher the probability of success of the attack. Of course, there is a practical obstacle since a video of a particular person, possibly required to perform liveness check movements is required to be replayed through the screen, which usually cannot be easily obtained (see Annex B, Figure 27).

3D photo/video render replay. This consists of reconstructing a static 3D face image or 3D face video model from a static face image reference and presenting it through a screen to the camera performing the RIDP process (see Annex B, Figure 28).

3.1.1.7 Face morph

Face morphing is a digital image processing technique performed by fusing two face images to form a synthetic face image that contains characteristics of both source faces. The goal of a morphing ([16]) attack is to create a synthetic face that can match the biometric templates of both individuals whose facial features were used to create it. Face morphing attacks can be performed both in presentation and injection attacks (see Annex B, Figure 29).

3.1.1.8 Deepfake Replay

As mentioned earlier in this report, deepfakes are deep learning-powered software which are capable of generating synthetic photos and videos, realistically representing persons who never existed or movements or spoken expressions that were never performed. Based on the algorithms, training datasets and additional pixel blending and enhancement techniques, the results are so realistic that they are hard to distinguish from natural, legitimate content (see Annex B, Figure 30).

In the context of presentation attacks, once a deepfake is properly generated and programmed to perform the required movements or expressions, it is replayed through a screen (TV,

([16]) The term morphing attack, along with the deriving taxonomies and related detection techniques (MAD) is a discrete topic, defined and examined mainly in the contexts of Automatic Border Control (ABC) and Electronic Machine-Readable Travel Documents (eMRTD) enrolment. In this document, face morphing is considered an attack instrument used in presentation and injection attacks against an RIDP system. Thus, in this context, PAD and IAD controls are considered applicable to morphing attacks, and further analysis focused exclusively on MAD is omitted. For more information on MAD, please see https://www.christoph-busch.de/projects-mad.html and https://arxiv.org/abs/2011.02045.

computer, tablet, mobile phone) against a camera that performs the RIDP process (see Annex B, Figure 31).

Deepfakes are described with more detail as a core attack instrument of digital injection attacks in Chapter 3.2.2.

As mentioned earlier in this chapter, impersonation is not the sole objective of adversaries; identity obfuscation is also an objective. Some of the common obfuscation attacks are the following.

- **Extreme makeup/disguise.** An effort to disguise face characteristics so that face recognition does not detect mismatches.
- **Partial occlusion.** The use of glasses, hats, clothes or other items, with the purpose of limiting visibility and evaluation of physical characteristics.
- **Bad quality of video or audio.** A deliberate attempt to have bad video or audio quality in order to bypass the RIDP platform's controls.

3.1.3 Attack Methods

Considering the description of presentation attacks and summarising the use of relative attack instruments, the attack methods used are described below.

Biometric Artefacts

The aim is to present a fake manufactured biometric in order to circumvent the security controls of the system. This can be conducted with the use of a mask imitating the face of the legitimate user. The mask can be manufactured from several types of material in an attempt to be as close as possible to the live physical user. Materials appropriate for this type of attack vary: 3D and 2D printed masks, silicon, printed texture, etc. The most successful instruments for this attack method are advanced silicon masks, which can feature details like skin texture and other distinct characteristics of a live face. Moreover, 3D modelling can contribute to making this attack method more successful in some cases. Liveness detection controls normally have the ability to adequately detect and block these attempts. Fake biometrics can also be created with cosmetic surgery or makeup methods. In this case, it would be more difficult for the attack to be systemically detected and may require the intervention of a human agent.

Photograph of the Biometric

A photograph or a digital image of the real user can be presented instead of the actual biometric (face) to the face recognition system which attempts to match it with the photograph of the ID document. The quality of the photograph can be of high level. In this case, controls like depth-sensing cameras or sensors or texture analysis (advanced systems analyse the texture and micro-movements of the skin, such as pulsations, to identify signs of liveness) can be used to detect photographs. Moreover, if the systemic controls are not adequate, a human agent's check during the real-time session will easily foil this attack.

Replay of a Fake or Recorded Video

The video can be used as attack method, either by creating a fake video or by capturing a video of the real user and replaying it. This attack method has the advantage that it may be closer to the real person performing the required biometric action, such as facial recognition or a specific gesture. The challenge–response method can commonly address this method, if the movement which is requested is not the same each time and has not been recorded or cannot be reproduced appropriately. Temporal analysis (which analyses the temporal characteristics of the presented data, looking for patterns that indicate a live action versus a static recording) can also be important to make the remote ID system relating to this attack method more robust.

3.2 INJECTION ATTACKS

3.2.1 Overview

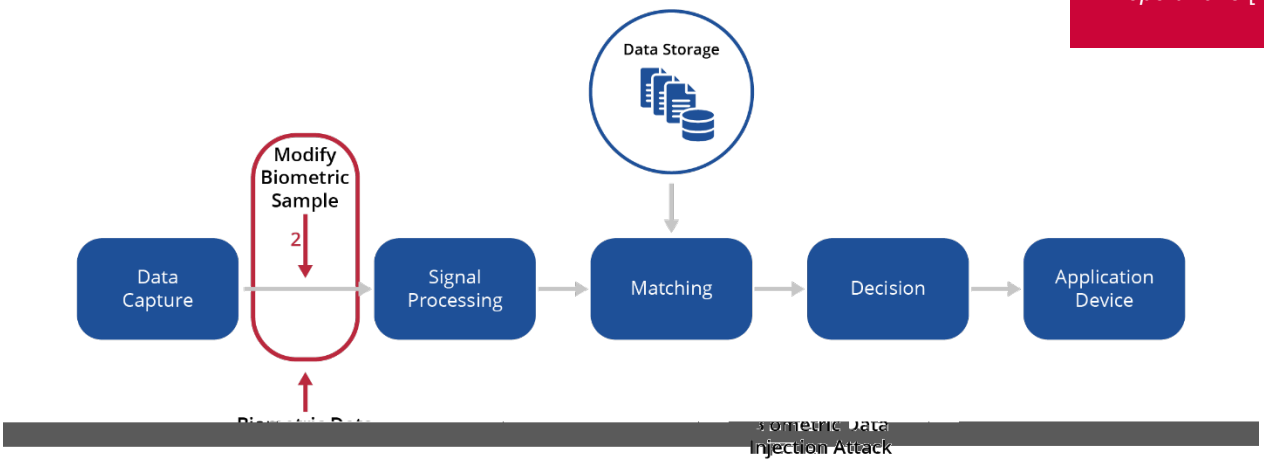
Biometric data injection is a type of **circumvention attack**. It falls under the types of attacks against the biometric but differs from a presentation attack, since for its execution it does not rely on the direct exposure of an attack instrument against biometrics capture element, but rather **by generating an artificial, digital biometric artefact which is injected directly to the biometrics recognition subsystem. In this way, it aims to bypass the PAD controls enforced during the biometrics presentation, and thus the biometrics software receives fraudulent input prior to performing detection and recognition**. The following diagram depicts the exact point in a biometrics system where data injection attacks take place. The difference in relation to the presentation attack point is also shown.

INJECTION ATTACKS
Five times more frequent and sophisticated than presentation attacks*

**as observed through iProov's iSOC global operations center.*

Source: iProov Threat Biometric Intelligence Report 2023 [15].

Figure 6: The point where biometric data injection attacks take place



The emergence of injection attacks has already been observed ([17]) by the various stakeholders of biometrics and identity proofing ecosystem, and ENISA had already identified that video injection is the most promising attack method in its 2022 report.

Digital injection attack incidents surged during 2022, with approximately five times more frequent and sophisticated incidents than current presentation attacks, and a further increase is expected in the following years. Factors driving this rapid increase are:

- the nature of the attack, which allows scaling, automation and lesser involvement by the threat actor in comparison to presentation attacks;
- public availability of synthetic imagery generation tools that are capable of designing and launching digital injection attacks against RIDP services;
- computer and GPU-accelerated recourses are easily available through cloud service providers (Infrastructure as a Service);
- Crime-as-a-Service ([18]) or the commercialisation of cyber criminality, which provides an easier way for lesser-skilled cybercriminals to achieve cybercrime activities, without relying on their own, limited or non-existing cybercrime skills and knowledge.

([17]) https://www.biometricupdate.com/202304/surprise-deepfake-fraud-on-the-rise.
([18]) https://enlets.eu/wp-content/uploads/2022/09/CaaS-1.pdf.

Impostors who choose to perform such attacks usually aim to impersonate a legitimate identity by matching their spoofed identity to one already known to the RIDP system.

The attacker needs to have control of the device which participates in the RIDP process, since specific technical preparatory steps are required for the execution of each injection attack type. This means that the device used to perform the attack is unsupervised.

Digital injection attacks have a greater success rate than presentation attacks, due to two main reasons:

- the nature and complexity of the attack;
- the lack of mature and standardised detection methods and standards.

Although various methods are proposed for this matter in the bibliography, unlike PAD, no globally accepted standards currently exist. CEN TC 224/WG 18 is working on an upcoming standard on injection attacks detection, but it is not expected to be published before the end of 2024.

Before elaborating on the various types of injection attacks, a brief overview of the basic elements of injection attack are presented.

The basic elements of an injection attack are:

- the injection attack instrument (IAI);
- the injection attack method (IAM).

By the term injection attack instrument, we consider a digital representation of the biometric, in the form of a recorded, synthetic or live photo or video. The biometric can be genuine or modified using a variety of ways, such as deepfake, face morph, face swap or recorded replay.

By the term injection attack method, we consider the specific methodology to interfere with the RIDP system and modify or replace the original biometric data captured, before being transferred to the RIDP service, so that fraudulent submission is possible.

In an attempt to provide a general overview of the injection attack types, the section presents a brief overview of the common injection attack methods and instruments. In all cases, it is considered that the RIDP process involves a user device (mobile or computer) running a web browser or local application. For examples/demonstrations of the various attack instruments, please consult Annex B.

3.2.2 Attack Instruments

3.1.1.9 Replay

During a replay attack, a recording of a human performing a previous authentication action is injected during the RIDP process, bypassing the biometric capture. The difference between this method and the rest of the injection attack types is that the content is genuine (not synthetic), but corresponds to a previous point in time and may or may not reflect the identity which is currently performing the authentication attempt.

3.1.1.10 Face morph

As stated in the previous section of PAIs, face morphing is also a form of injection attack and is indeed a quite frequent attack instrument chosen by adversaries. Note that the results of face morphing are generally more realistic than the relatively simplistic method of face swaps, due to

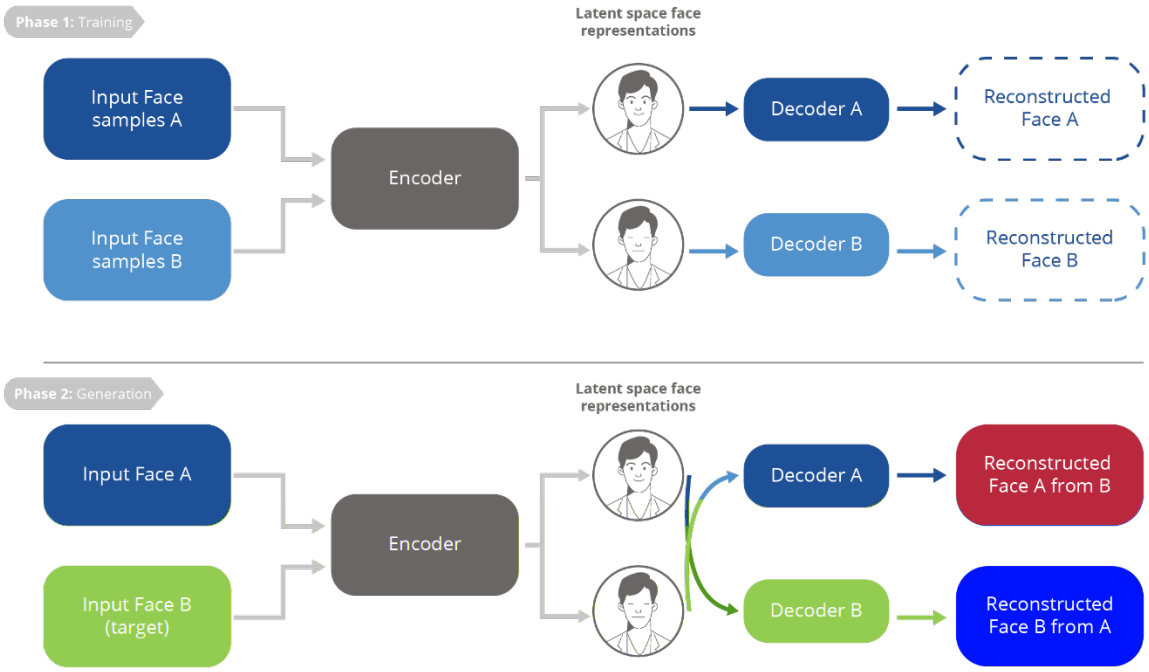
the underlying pixel blending mechanisms involved in the technique, which can be improved even further with neural morph enhancement techniques (see Figure 32).

3.1.1.11 Deepfake

Deepfake ([19]) is a technology utilising the power of deep learning to produce audio and visual content. The underlying technologies that contribute to the emergence of this kind of fraudulent content are autoencoders and GANs. Deep learning is a type of machine learning, mimicking the human brain's neural networks for pattern identification, which in turn achieves generation of convincing audiovisual results.

The first popular category of deepfake generation is the autoencoder. It is a neural network designed to learn essential data from a training set, (e.g. face images) through a process of encoding, compression and decoding, which ultimately leads to the reconstruction of the given data. Autoencoders area applied in various sectors, such as computer vision, facial recognition and handwriting analysis. The diagram below depicts a simplified, high-level representation of an autoencoder model.

Figure 7: High-level model of an autoencoder



The two basic phases are depicted: the training phase, where the network is trained to detect facial attributes, and the generation phase, where the network reconstructs a synthetic face, having as input a source and target face. The neural network models currently used are capable of working with video samples and have several implementation variations, such as the variational autoencoder (VAE) or long short-term memory ([20]). These details fall outside of the scope of this report. However, readers are encouraged to consult the bibliography sources at the end of this document.

([19]) The term, a combination of 'deep learning' and 'fake', is thought to be attributed to the nickname of a user in a popular internet forum where, in 2017, started posting fake celebrity videos. Although the threads and the subcommunity were banned, the trend had been established among the community.
([20]) https://insights.sei.cmu.edu/blog/how-easy-is-it-to-make-and-detect-a-deepfake.

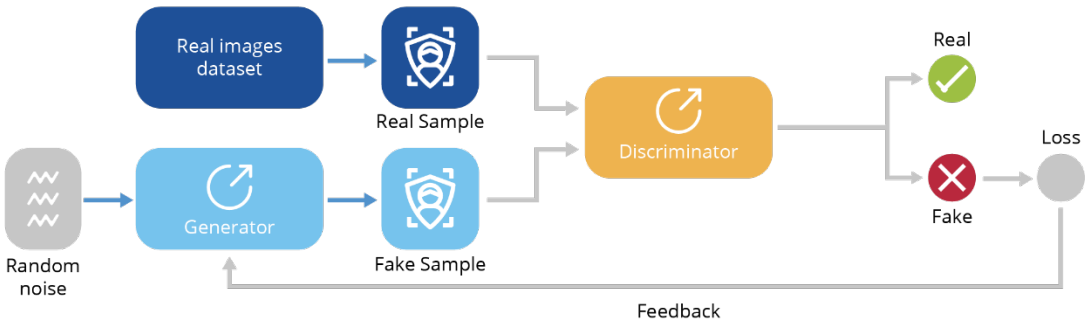
The results of autoencoder-based deepfake software are frequently very convincing ([21]), portraying a photo or a video of a person's face, performing movements or speaking in ways they would not normally do or that never existed in the first place. This is mainly due to the fact that autoencoders perform remarkably well in eliminating visual noise, while retaining the valuable information for the purposes of learning/decoding. This leads to the production of efficient models that can produce satisfying results, even when the given content is significantly different to the original training dataset.

GANs are the next step in deep learning-based synthesis, where a self-improving trained model can generate audiovisual results that are even more realistic, convincing and harder to differentiate from legitimate subjects. It can also maintain liveness features and thus may go undetected by traditional anti-spoofing methods. Face swapping is a popular application of deepfake technology.

The concept is based on two competing networks: the **generator** and the **discriminator**. While the first tries to generate results based on the input dataset, the latter provides feedback against the results generated. Through this feedback loop, the generator network continually improves and both networks gradually converge, resulting in a GAN capable of producing realistic results.

Training datasets are required as an input to produce satisfactory results, based on various manipulation techniques such as entire face synthesis, facial attribute manipulation, identity swap or expression swap. Such datasets are becoming increasingly available on the internet, helping threat actors to produce quicker, high-quality, high-success fraudulent content (see Figure 33).

Figure 8: High-level model of a GAN-based deepfake generation model



The various deepfake subtypes (e.g. face re-enactment, face replacement, face editing, face synthesis) serve specific attack models and depend on the particular type of neural network utilised, each one designed to work with a specific type of source identity sample and human visual input. However, GAN approaches are not seen solely as standalone tools, but are also incorporated in hybrid approaches, where one or more GAN models are combined with CGI methods [19], which shows the multitude of ways to generate such content, albeit with varying levels of quality and capability.

Today, a perfect approach to facial synthesis does not exist and face video synthesis remains a more complex matter in relation to photo synthesis. While autoencoders and GANs have their own set of limitations and research on the topic is proceeding at a high pace ([22]), efforts to

([21]) For realistic demonstrations of face morphs, see https://this-person-does-not-exist.com.
([22]) https://blog.metaphysic.ai/the-future-of-generative-adversarial-networks-in-deepfakes.

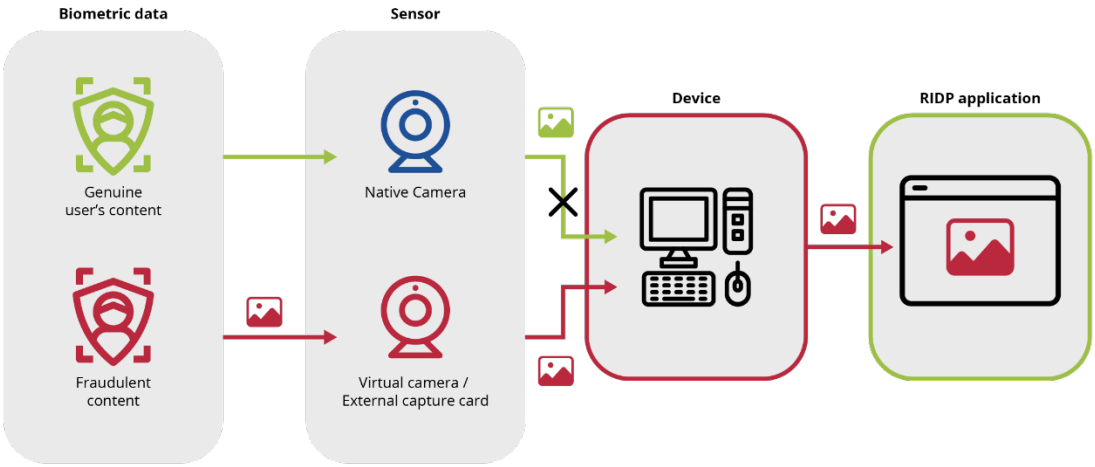
defend against the weaponisation of these approaches should intensify, following a proactive rather than reactive mindset.

3.2.3 Attack Methods

This section showcases the common IAMs against RIDP systems. Please note that the attack methods listed below can be combined in a final attack vector. This should be taken into consideration when performing risk analyses relating to biometric injection attacks, in order to adjust any corresponding security measures accordingly.

Virtual camera. This is the simplest case, where a virtual camera is introduced in the attacker's device by installing a programme that creates a software-based camera device within the operating system. A virtual camera allows the user to provide audiovisual content to applications, replacing the physical camera's input with multimedia files residing locally.

Figure 9: The concept of a virtual camera injection attack method

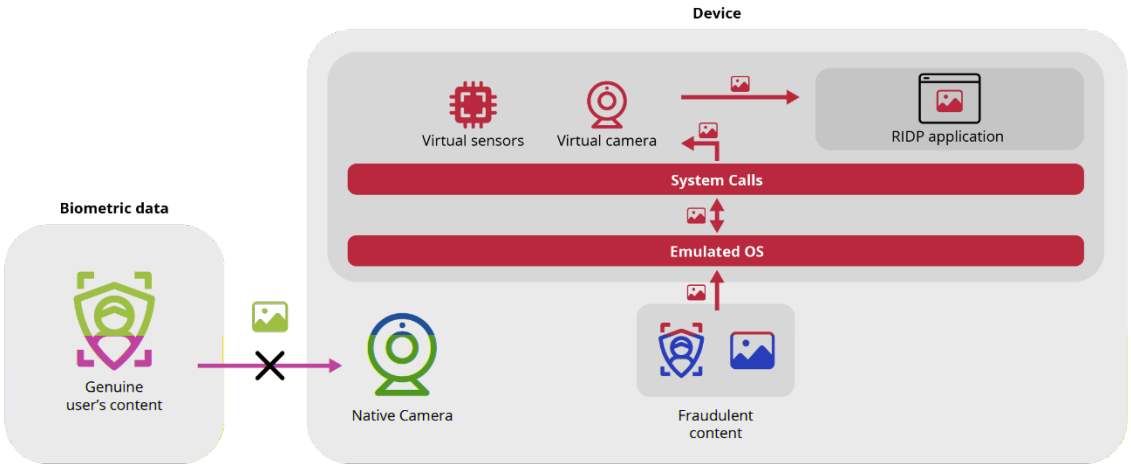


A variation of this concept is to use a physical, external video capture card, capable of accepting various video sources as input (e.g. computer output through HDMI), and presenting the final video stream to the user's device (e.g. computer or mobile phone) as originating from the device's native camera (see Figure 34).

It should be noted that virtual cameras are not considered an offensive type of software per se. They are freely available as a convenient utility software used in various cases, such as to provide video input to multimedia applications if the computer does not have a physical camera attached.

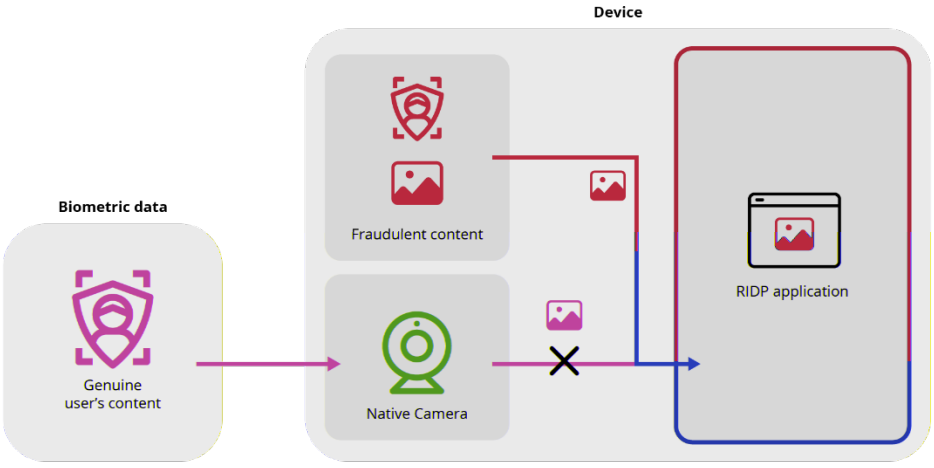
Device emulator. In this case, a computer with a mobile device emulator software is used, creating a virtual, emulated operating environment of a smartphone and its specific operating system. By using an emulator, the attacker can have access to all system calls being made by the RIDP application or web browser, and modify signals from the various simulated sensors (e.g. accelerometer, GPS, camera, microphone).

Figure 10: The concept of a device emulator



Function hooking. This concept is based on altering the original flow of system calls during runtime so that fraudulent content is injected during the biometrics capture procedure, replacing or altering the original biometrics that have been or will be captured. At a high level, the attack is realised as follows.

Figure 11: The concept of the function hooking injection attack method



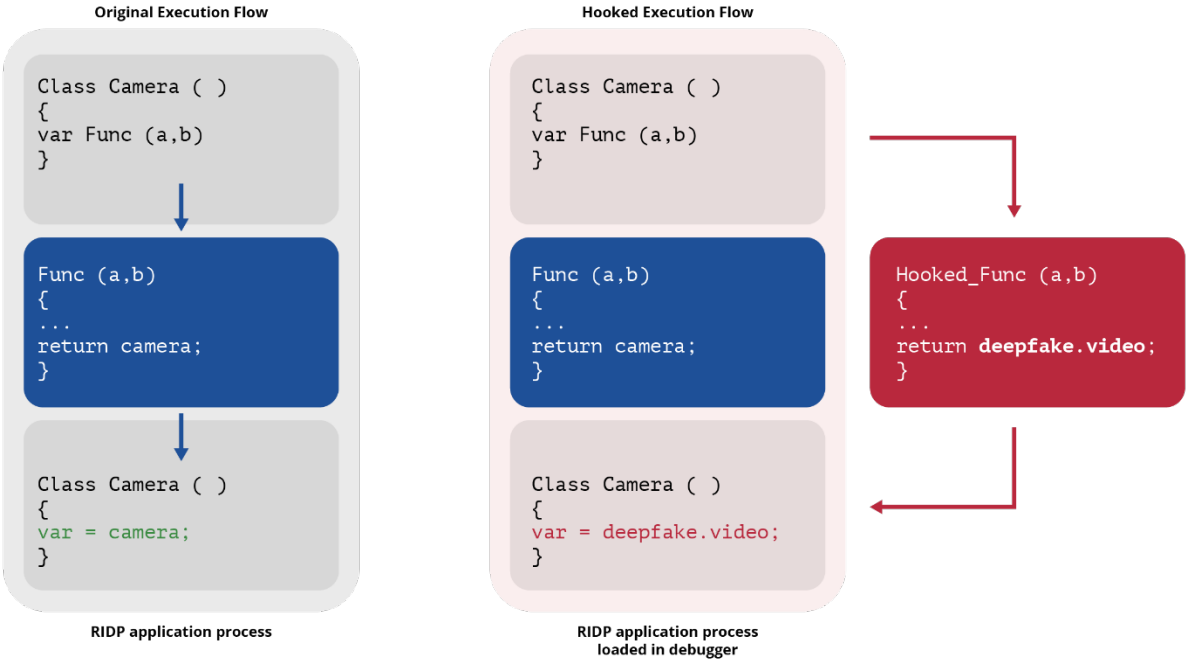
- The client-side environment is prepared by rooting the smartphone device and installing a custom ROM, in order to be able to debug in real-time the execution of the RIDP application or the web browser. There are also advanced debuggers which offer the same capability without the need of rooting the smartphone.
- Next, the RIDP application or web browser requires decompilation, in order to find the functions and the related code which call the device's embedded camera, in order to capture the photo or video for the needs of RIDP.
- Additionally, the attacker needs to write a function which will be executed when derailing from the main execution. The function will generally read the fraudulent multimedia file residing

locally (e.g. deepfake video) and then continue the normal execution of the RIDP application or web browser.

- The RIDP application or web browser process is loaded in the debugger installed previously.
- The RIDP process is initiated and reaches the point right before connecting to the device's camera for taking a photo or video.
- At that point, the debugger instructs the RIDP application or web browser to execute the custom function rather than the original, native function of the system.
- The custom function is programmed to read a local multimedia file instead of the camera's sensor and pass its contents to the next phases of execution.
- Once the file is read, the execution is continued with the normal flow.

In that way, the RIDP application is not aware of the circumvention achieved and forwards the injected, fraudulent multimedia content to the RIDP service.

Figure 12: Original execution flow of an application process versus a hooked flow, with the help of a process debugger



The complexity of this attack's implementation depends on the underlying technology used, for example a computer or mobile phone, operating system and particular application. Generally, this requires hands-on expertise and knowledge of specific technical procedures relating to software reverse engineering, debugging, device rooting and scripting.

Of course, an injection attack's success is not solely dependent on the attack method, but also relies heavily on the quality of the actual attack instrument.

Man-in-the-middle. In this case, the attack takes place at the network level and not at the application level, as in the previous cases. It is a network-type of attack, not exclusively applicable to RIDP or biometrics, but usable in any network topology, as long as the attacker has the method and tools to intercept the network traffic exchanged between two communicating parties. During an attack, exchanged messages can be intercepted, read, modified and relayed, while the legitimate participants are not aware of their confidentiality and

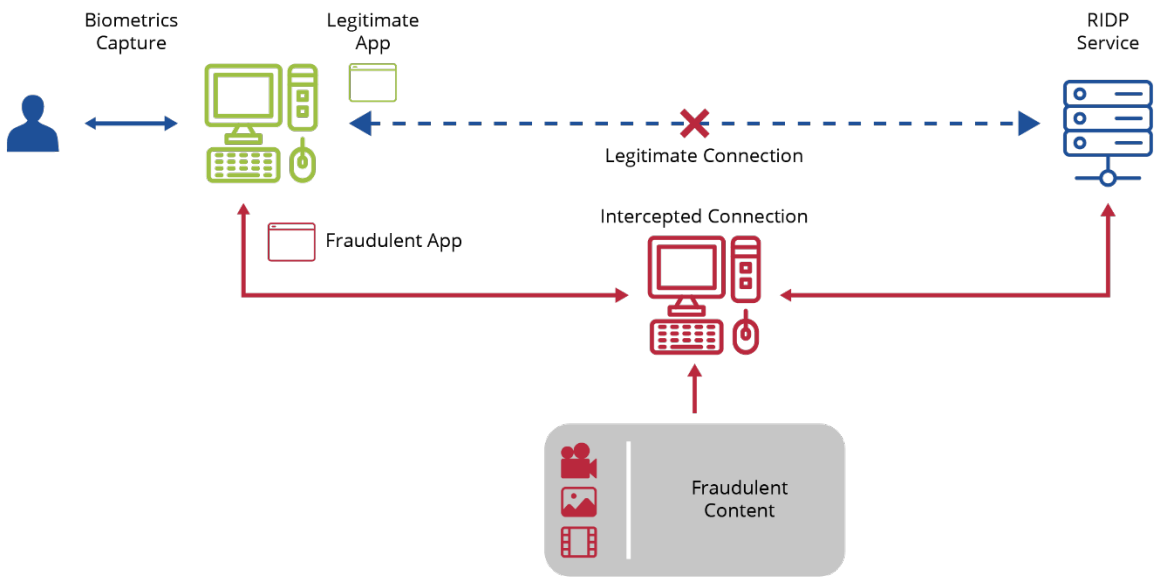
integrity breach. Generally, these attacks are possible through traffic interception or traffic redirection.

In the case of traffic interception, the attacker uses specialised interception software, installed in the victim's device or by eavesdropping on the local Wi-Fi or cable network.

In the case of traffic redirection, various techniques exist depending on the OSI level. For example, on local networks, IPv4 ARP spoofing, IPv6 router advertisement or automatic proxy discovery can be exploited, while at the internet level, DNS spoofing, DNS record/domain takeover or even BGP hijacking are widely used to point legitimate hostnames to fake servers.

It is also possible to perform such an attack with a fraudulent, spoofed mobile or web application which mimics the name, branding and general look and feel of the legitimate application. In that way, it aims to trick the user to install it. The application will then act as a proxy which allows the adversary to intercept the user's traffic. This method assumes that applications of particular RIDP providers would be targeted.

Figure 13: The concept of the man-in-the-middle injection attack method



It is obvious that multiple combinations can be used to craft and conduct a biometric injection attack, making the topic of defence against injection attacks quite complex. This does not mean that all combinations are equally successful, since all are subject to multiple complexity factors, making the final attack more or less feasible, depending of course on the benefit that the adversary may have from a successful attack. These complexity factors describe the time and expertise that the attacker must invest, spanning from the initial planning phase of the attack to the exploitation phase (technical implementation and the actual execution). The following indicative factors are considered:

- the level of information research regarding the targeted RIDP system (underlying technology, method, workflow);
- the equipment (software/hardware) that may be required to conduct the attack;
- the level of access to biometric sources of the victim;
- the particular RIDP method regarding the level of involvement of a human operator during the process (e.g. human operator actively observing versus a fully automatic method).

For assessing an RIDP solution's performance on biometric data injection attacks, the upcoming CEN TC 224 standard on this topic is expected to include an evaluation framework built around those complexity factors.

3.3 IDENTITY DOCUMENT ATTACKS

3.3.1 Overview

During an RIDP session, apart from the face biometric capture, identity documents are also used as a means of verifying an identity and are thus frequently attacked. In fact, identity document attacks have seen a significant increase within the last year in some Member States, with approximately 9 out of 10 identity spoofing attack attempts relating to identity documents.

There are numerous valid identity document types either in traditional (paper-based) or electronic form. This variety poses a serious challenge during the RIDP process, not only regarding the technologies that must be utilised to acquire the attributes of these documents, but also the complexity in maintaining procedures and information for inspecting these documents in a manual, human way, which is still frequently the case.

Prior to presenting the identity document attack landscape, a brief overview of the various identity document types and their features is presented in the next paragraphs. Readers can consult ENISA's 2022 report for a more detailed view on identity document types and features.

Based on the security technology used, identity documents can be either traditional (paper-based) or electronic.

- **Traditional, paper-based documents.** These are the oldest and simplest type and cause a multitude of challenges, as most of them were not designed with remote identification in mind. The level of assurance they can provide is proportionate to the number and quality of the security features applied.
- **Electronic identity documents.** These use an embedded electronic microprocessor chip which contains biometric information that can authenticate the identity of the holder. The identity information is printed on the document, repeated on the machine-readable zone (MRZ) section and stored in the electronic chip. Public key infrastructure (PKI) is used to authenticate the data stored in the chip, making it impossible to forge with current technologies when all security mechanisms are fully and correctly implemented. A transition to post-quantum cryptography is expected to be required to protect from quantum technology capable of breaking today´s encryption algorithms.

The main types of identity documents with the relative characteristics that allow the acquisition of their data can be summarised as follows:

- traditional paper-based;
- MRZ-based;
- Contact chip-based;
- NFC-based.

It should be noted that an identity document is usually a combination of the above types and incorporates multiple security features, such as printed visual security features, an MRZ and a contact chip.

Identity document attacks aim to spoof or obfuscate an identity by attempting to achieve a match between the fraudulent face represented during the biometric capture and the face depicted in the identity document. In general, an attack based on forging an existing, legitimate identity document part (e.g. photo or textual identity information) has a higher success rate than

crafting a completely fictional identity. INTERPOL ([23]) provides the following categorisation of identity document fraud.

False documents:

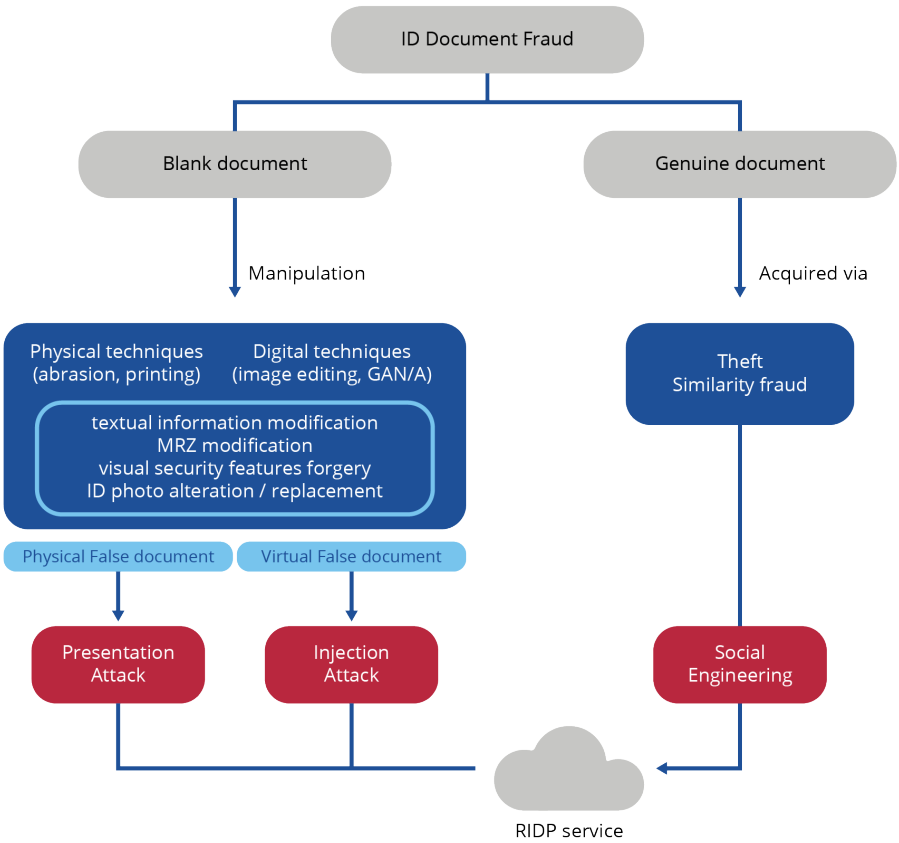
- counterfeits: unauthorised reproductions of genuine documents;
- forgeries: alterations of genuine documents;
- pseudo documents: documents replicating codes from official documents (e.g. passports, national identity cards) but not officially recognised.

Genuine documents:

- genuine documents obtained through fraudulent activities (e.g. theft, robbery, blackmail);
- similarity fraud, where genuine documents are misused by an impostor, using the personal information of someone who shares physical or behavioural similarities with the targeted (victim) identity.

Based on the forgery techniques and the methods carried out to submit the fraudulent document to an RIDP service, identity document attacks can be considered either presentation or injection attacks. It is therefore obvious that similar PAD and IAD methods should be developed and applied for identity documents. A pain point on this matter, as confirmed by a number of stakeholders, is that the maturity of PAD and IAD methods applicable to identity documents is somewhat lower than those applicable to biometric attacks of the human face.

Figure 14: The types of identity document fraud in attacks to an RIDP service



([23]) https://www.interpol.int/en/Crimes/Counterfeit-currency-and-security-documents/Identity-and-travel-document-fraud.

The latest example ([24]) of such a presentation attack on identity documents is the Computer Chaos Club Video-Ident attack which, based on the analysis presented, circumvented the existing controls for online, video-based identification and managed to access the personal health record of a test person.

Other similar, recent examples and demonstrations show the necessity of prioritising the research and development of adequate and effective technical and procedural controls focused on identity document attacks, following a risk-based approach and including harmonisation across the EU, since several gaps regarding the permitted RIDP methods are already evident among Member States.

Responses from the survey and the interviews highlighted that the threats against identification documents remain quite concerning. Although NFC-reading of identity documents significantly reduces the likelihood of a successful attack (where possible), there are still identity documents without electronic elements in circulation. Although these types of documents incorporate security elements such as printed holograms, UV areas, detailed and complex printed patterns (e.g. guilloché) and MRZ, it is easy to obtain a fake document with these elements or even create a fake digital or physical form of such a document. Another obstacle is that testing of solutions capable of detecting document forgery is very limited, since in most European countries, the creation of forged identity documents for testing purposes is forbidden by law. Testing can still be made in a few cases, but only after approval and under surveillance by the police. Additionally, it is perceived that ID document attacks utilising information and characteristics of an existing person's identity have higher chances to succeed than attacks relying on artificially crafted identity information of a non-existing person. Finally, the idea of maintaining a database containing the latest versions of official identification documents issued across the EU, accessible in a controlled way by TSPs and RIDP providers, has been expressed by some stakeholders. This could provide further assurance in the identity proofing process performed by these entities.

([24]) https://www.ccc.de/system/uploads/329/original/Angriff_auf_Video-Ident_v1.2.pdf.

4. GOOD PRACTICES

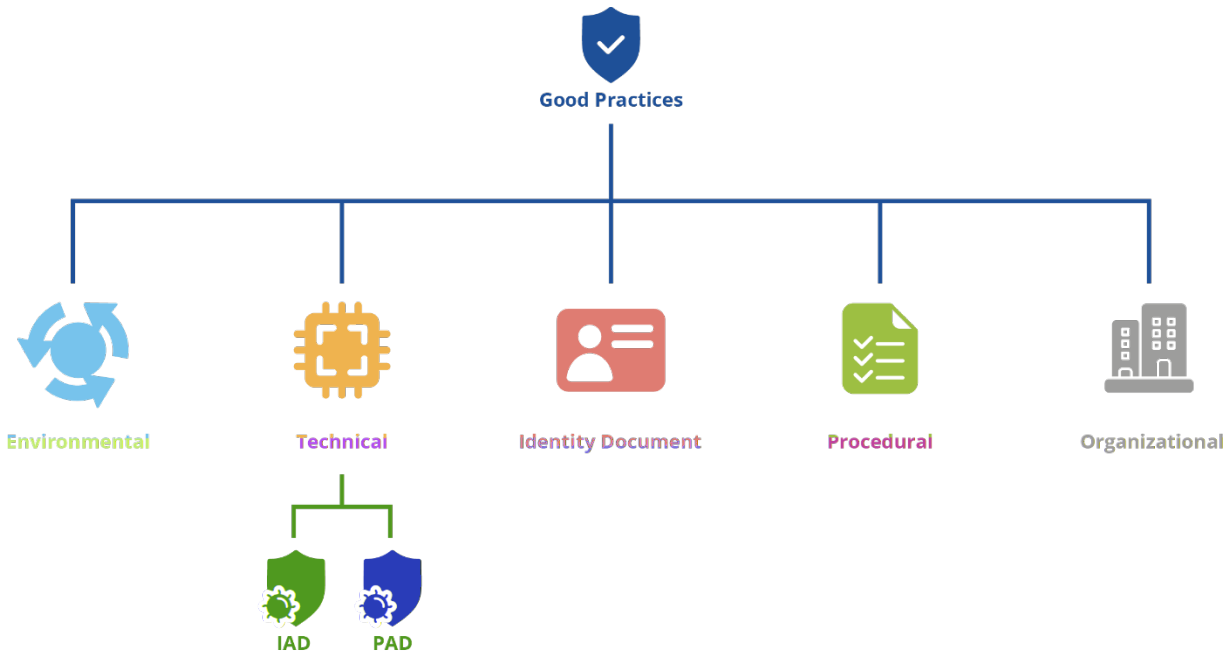
The various attacks described in the previous chapter can pose a complex and multi-dimensional problem for RIDP techniques. This chapter focuses on presenting an overview of the various countermeasures identified and analysed during the bibliographic research phase and the surveys and interviews of the various stakeholders.

Since a silver bullet for the defense of an RIDP system does not exist, a multi-layer, risk-based approach is recommended, where all individual security layers contribute to the overall security of the system, taking into consideration the intended use-cases and the level of assurance required. As in most engineering problems, the selection of countermeasures should aim for the right balance between effectiveness and usability.

The various good practices presented in this chapter can be divided into the following categories:

- environmental controls;
- technical controls, consisting of:
  - PAD controls,
  - IAD controls;
- identity document controls;
- organisational controls;
- procedural controls.

Figure 15: Categorisation of good practices



4.1 ENVIRONMENTAL CONTROLS

Control of the environmental conditions in which the RIDP process takes place is essential, since the information retrieved can be analysed and provide preliminary information regarding the existence of synthetic identity or other indicators which may be correlated with a fraud attempt. The proposed environmental controls are described below.

Proper lighting conditions. These ensure that information is captured more accurately, facilitating the optimal execution of identity proofing. Low-light conditions and strong backlight tend to produce noisy and pixel-saturated photos and videos which are more difficult to analyse by the various algorithms.

Definition of minimum multimedia specifications. To guarantee the quality of the evidence, it is important to define the minimum acceptable photo and video criteria, such as the post-compression video bitrate and photo/video resolution, along with the existence of a physical microphone device. Additionally, since deepfake generation tools tend to produce videos with high bitrate, this indicator could be taken into account in the deepfake detection controls incorporated by the RIDP solution. Another way to detect deepfake attacks is based on the analysis of the frames per second (FPS) rate of the video. In a real video stream captured by a camera and streamed in real time to the RIDP service, the FPS rate experiences small fluctuations, which is an expected effect when capturing and transmitting live video. A video stream with a constant FPS rate could be considered an indicator of a possibly fraudulent video rendered by a deepfake generation tool.

RIDP application client-side architecture. In a client-server RIDP system, the first vulnerable element is the user's device; thus, the enforcement of technical controls should cover both the client and server sides. A dedicated client-side RIDP application would offer greater flexibility and more effective and granular client-side control enforcement than a web application, where the RIDP process involves solely a web page and a web browser running in a user's device.

Regarding the implementation options of RIDP solutions, as analysed during the information gathering phase of the report, two main models have been observed: application programming interfaces and software development kits. Although the selection between the two will depend on the intended use case, the latter option is considered to provide a somewhat increased level of security, mainly due to the following reasons:

- ability for more granular control enforcement and signals/information collection in comparison with a web app;
- adds a layer of complexity through code obfuscation and runtime execution protection environment (if available, although an attacker with the required skills can overcome it).

4.2 PAD CONTROLS

Presentation attacks are the first type of biometric attack and, although injection attacks seem to have exceeded the level of sophistication and frequency compared to presentation attacks, technological evolution – especially GANs and AI – provides continuous benefits in combating both attack types. There are numerous PAD approaches, thanks to ongoing worldwide research. Some commonly used approaches are described below.

Hardware-based, which exploit camera characteristics like variable focusing properties, degree of depth or effect of defocus; these methods are relatively efficient as they do not involve any additional device besides the original camera.

Software-based, which consist of static, dynamic and neural network / deep learning-based approaches.

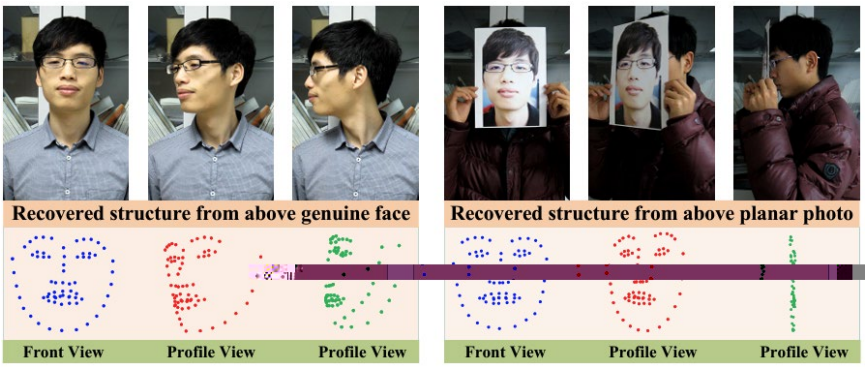
Static approaches aim to detect frequency, image quality and texture-based artefacts by analysing the micro-texture of the surface presented to the camera. In general, these methods can detect photo and video replay attacks but not high-quality 3D mask attacks, and generally require minimal computational resources.

Dynamic approaches aim to detect facial texture or face, pupil and body motion-related artefacts. They can address photo attacks and some kinds of video replay attacks (low quality but not sophisticated deepfake attacks) and require more computational resources.

In recent years, the emergence of neural networks has allowed newer ([25]) techniques to be developed, aiming to extract deep-level features automatically. Examples of these modern types of PAD controls are shown below.

3D geometry-based, which analyse the three-dimensional depth and geometry of a face and produce a 3D representation (3D map) of a face's features, taking into account the natural, perspective distortion based on the capturing distance. With 3D liveness technology, the number of variables and involuntary, 3D liveness signals are significantly increased compared to 2D methods. This makes 3D liveness solutions more reliable in terms of accuracy, usability and attack preventability.

Figure 16: A comparison of recovered sparse 3D facial structures between genuine and photo faces, showing significant differences [6]



Please note that in the above example, the sparse 3D facial structure is reconstructed from a subject located at a fixed distance from the camera. Another approach would be to reconstruct a 3D representation of the face based on the perspective distortion that occurs between multiple face shots, as the subject moves closer to the camera.

Phoneme-viseme mismatch detection, mainly focusing on lip-sync deepfakes, aiming to detect inconsistencies between visemes (dynamics of the mouth shape) and spoken phonemes. The technique works against a common flaw of current AI technology, which has imperfections when trying to visually match mouth movements with spoken words.

Light absorption-reflection analysis, aiming to remotely estimate the heart rate of the face video through light absorption by blood cells or microvariations in the skin colour caused by light reflection and spot anomalies, which could indicate fraudulent content. A similar approach [17] is related to the illumination of the screen's light to the user's face skin.

The above examples of neural networks and machine learning-based PAD software require training through datasets, so that content is analysed and classified, leading to the detection of fraudulent content. A core challenge of these PAD is the need for continuous maintenance and

([25]) https://www.unite.ai/best-deepfake-detector-tools-and-techniques.

update of the datasets, so that the PAD software can detect attack novelties, while the validity and consistency with the purpose of the algorithm is always ensured.

Combined PAD methods: simultaneous use of multiple technologies in the same system would be recommended where possible, provided it does not reduce usability or accessibility. Also, when combining biometric PAD methods, it should be ensured that all individual methods provide the same false acceptance rate; otherwise, the overall security and assurance provided is at stake.

4.3 IAD CONTROLS

The current gaps in standardised detection methods make it imperative for actions towards the evaluation and standardisation of IAD methods.

Also, one of the main challenges in effective protection against injection attacks is the rate of technological evolution, which is currently exploited by threat actors while defenders try to keep up.

Conceptually, IAD happens at the intersection of three domains:

- PAD;
- general synthetic image detection;
- cybersecurity protocols and techniques applied on the device and data exchange channels.

The recommended practices fall under the categories of **preventive** and **detective** controls.

Preventive controls

- **Camera anti-tampering.** This aims to secure the communication path between the device's native camera and the RIDP application or web browser, ensuring the authenticity and fidelity of the captured content. This particular topic is under ongoing research and development, with various technical approaches currently published (e.g. media content authenticity provenance ([26]), cryptographic image attestations [7], zero-knowledge proof-based image attestations [8], camera identification through image hashing [9], photo authenticity through a mobile trusted execution environment [10] or steganography watermarking [20]), varying in maturity or applicability. However, the underlying technical aspects of this topic fall outside of the scope of this report.
- **Deterrent software controls.** These aim to introduce multiple levels of additional complexity in performing virtual camera or function hooking attacks against the RIDP application. Two main types of this control category are:
  - **code obfuscation**, which protects from reverse-engineering and breaching the integrity of the RIDP application,
  - **runtime protection of the application** (e.g. runtime application self-protection, trusted execution environment), which allows the detection of rooted devices and function hooking attempts and terminates the execution or limits the functionality of the RIDP application.

Detective controls

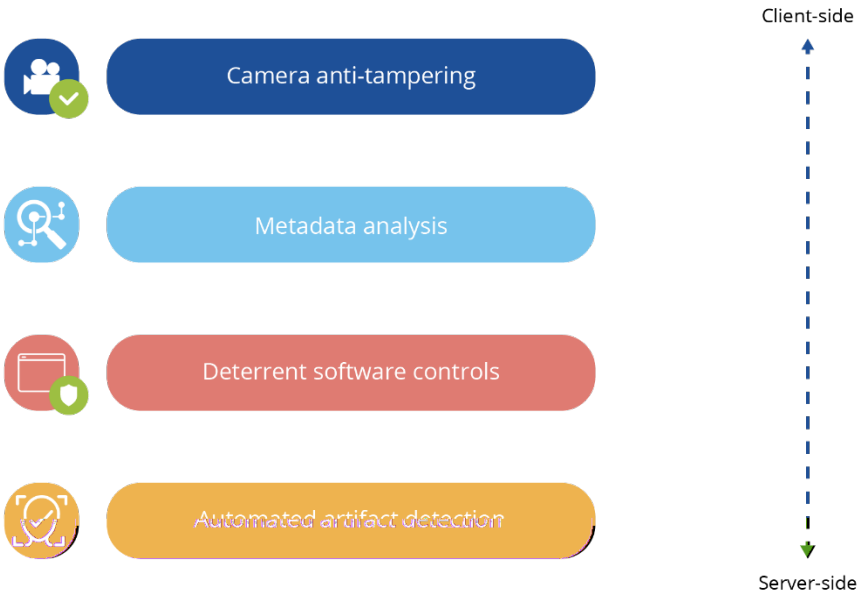
- **Session metadata analysis.** This aims to detect the existence of a virtual camera or device emulator in the user's operating system by examining various metadata of the RIDP session, such as the resolution of the photo/video stream, data from GPS, accelerometer, gyroscope, timestamps, network information as well as operating system and user-agent fingerprinting, etc.

([26]) Coalition for Content Provenance and Authenticity: https://c2pa.org

- **Automated artefact detection.** This aims to identify video modalities (e.g. data moshing effect), audio/video anomalies and correlated inconsistencies in the biometric representation of the user, which would indicate potentially fraudulent content. Mechanisms utilising various types of convolutional neural networks (CNNs, e.g. region-based CNNs, deep CNNs) and preference for 3D face liveness over 2D are recommended, considering the developments in quality of the deepfakes that can be produced. Deep learning techniques primarily developed for PAD are also applicable here, since the various characteristics of the underlying algorithms, focusing on facial, motion and textural features extraction, can work in a generic approach.

It is important to stress out that since no silver bullet exists, the optimal solution is a multi-layered incorporation of the various control types in an RIDP system in order to ensure the integrity of the biometric capture process. The following diagram depicts the overlay of the various injection attack-related preventive and detective controls across an RIDP system.

Figure 17: The overlay of various IAD technical controls in a client-server RIDP system



4.4 IDENTITY DOCUMENT CONTROLS

The detective and preventive controls proposed herein fall under the technologies of document authenticity checks, PAD and IAD, focusing on verifying the optical and electronic features of the document, detecting synthetic artefacts and ensuring the integrity of the RIDP process.

Combined PAD and IAD methods. A combination of PAD and IAD methods to ensure the integrity of the biometric capture phase of the RIDP process is recommended, taking into account the latest developments in the threat landscape. There are also dedicated tools for detecting document forgeries and various datasets which support the development and functionality of such tools. However, they are somewhat limited compared to the datasets used for deepfake generation. For a more detailed view on PAD and IAD, please consult Chapters 3.1 and 3.2 respectively.

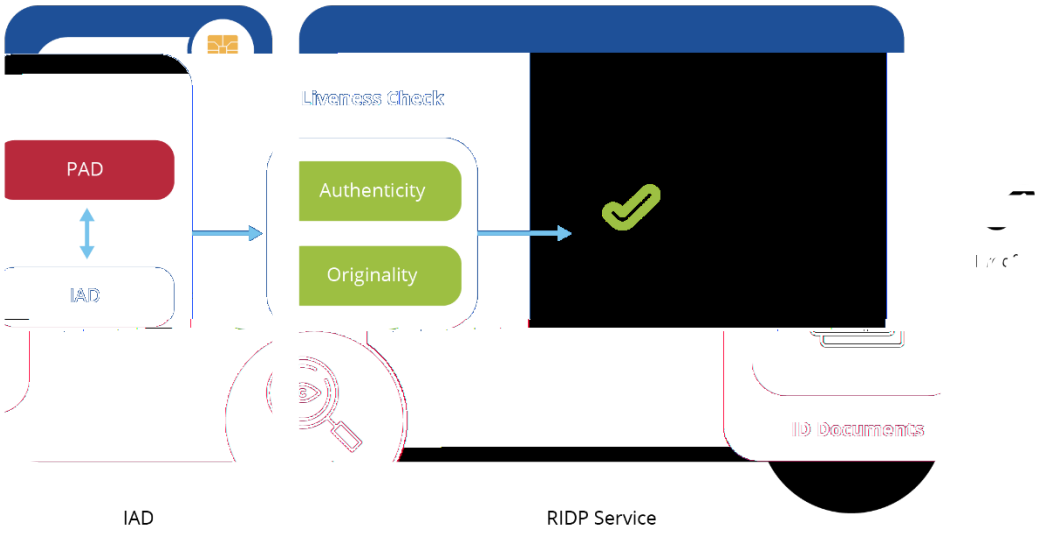
Liveness checks. These types of checks aim to prove the authenticity and originality of the captured document, by validating its physical structure and layout, information-encoding elements (barcodes, QR codes) and other possible, visual or electronic security features it

incorporates, and that the document instance presented during the RIDP process is the original and not a physical or digital copy. A combination of the following methods is recommended during the liveness check phase.

- **Visual security features verification.** The detection of properties of visual security elements is usually applied as a first level of defence, especially for documents that do not incorporate electronic security features. The existence of visual security elements (e.g. holograms, watermarks, guilloche prints, UV prints, micro-texts) along with some physical properties they have (e.g. reflections, gradient colouring) can be detected in photo or video RIDP sessions. However, the application of these features is not consistent across countries and among the various document types; thus, alteration is easier than the information included in an NFC chip. As a baseline, a minimum set of visual security features is always necessary to remotely verify the authenticity of an identity document.
- **OCR data extraction.** OCR software is used in real time to capture and analyse visual elements and subsequently extract data which is further processed to determine validity.
- **MRZ verification.** A process which validates the integrity of the data encoded in the respective area of the document, containing personal information of the holder in a standardised format.
- **NFC scanning.** This method is considered to provide the highest level of assurance since the biometric and subject information contained in the NFC chip can be cryptographically verified thanks to public key infrastructure technology and Card Verifiable Certificates. Additionally, the biometric digital photograph contained in the chip can be used for face matching with a higher accuracy level. Although technically possible, scanning the NFC chip of an official identification document to retrieve and verify the holder´s information and biometric face photo is not consistently permitted across the EU. This could potentially give attackers the opportunity to perform impersonation attacks with higher success, targeting Member States where NFC-reading by private TSPs and RIDP providers is not legally permitted. A relatively recent and interesting exception is Switzerland, where the Swiss Financial Market Supervisory Authority allows chip scanning in online identity proofing to enable companies to offer smooth onboarding for clients ([27]), along with appropriate measures for the verification of the authenticity and integrity of the information. Please note that even if reading the NFC chip is considered to provide the highest level of assurance, the achieved assurance level always depends on the level of security of the chip and its data.
- **Modern document security features.** New versions of national identity documents incorporate visual elements based on digital signatures, such as the QR code of the new French ([28]) national identity card, certifying the authenticity and the validity of the document.
- **Multimodal biometric verification.** Current travel document standards consider only the biometric data of the face as mandatory. Additional biometric data such as fingerprint or iris data are considered optional but could increase document security unless EU legislation on identity cards expands the permission to access such types of information beyond national and law enforcement authorities.

([27]) https://www.finma.ch/en/news/2021/05/20210517-mm-rs-16-07-online-identifizierung.
([28]) https://www.interieur.gouv.fr/actualites/actu-du-ministere/nouvelle-carte-nationale-didentite.

Figure 18: The general concept of technical controls for ID documents in RIDP



4.5 PROCEDURAL CONTROLS

Supplementing the technical controls described previously, there are procedural controls which contribute to strengthening the overall security of the RIDP process. These types of control are described below.

Challenge-response mechanisms, which aim to introduce complexity and non-predictability in preparing the fraudulent responses to be injected, since it will require a significant amount of time and effort to identify, prepare and inject them properly. As a rule, the higher the randomness of the challenges, the higher the complexity for the attacker. This control breaks down in two types, as shown below.

- **Active, high-entropy challenges,** requiring the user to perform verbal or visual, randomised actions from an extensive set of active challenges, so that they cannot be guessed by an attacker. Active challenges focus in detecting changes in face occlusion, face expression, view angle and ambiance. 'CAPTCHA' technology can also be utilised in conjunction with active, motion-based challenges, requiring the user to perform image, text or audio tests. An additional check can be the verification of a bank transaction between the user and the entity performing RIDP. Requiring the user to perform rapid head movements could allow the operator to spot anomalies in the video, since autoencoders and GANs are not quite capable of recreating diagonal face views, mainly due to lack of this type of information in the training datasets.
- **Passive, high-entropy challenges,** where the biometric capture system introduces visual changes in the user interface, (e.g. sudden blinking of an on-screen notification or display of chromatic sequence, overlay animation effects), aiming to detect an involuntary reaction of the user, reflections of colour onto the user's face, or inconsistency in the perception of the user's face and thus, verify liveness.

An important decision factor related to these high-entropy, randomised challenges is the balance between the complexity for the attacker and the accessibility and inclusivity levels provided for legitimate users with disabilities. Traditional, CAPTHCA-style mechanisms are not to be considered an effective challenge since they are already outperformed ([29]) by AI and that

([29]) https://www.theverge.com/2019/2/1/18205610/google-captcha-ai-robot-human-difficult-artificial-intelligence.

could possibly amplified, with AI being capable of mimicking human movements in the context of an active challenge.

A recent approach in challenge-response mechanisms, namely GOTCHA [11], is a combination of extensive active and passive challenges in a sequence of varying complexity, designed to detect artefacts introduced by modern deepfake generators.

The key concept of this approach is the cascading set of challenges. It is based on the assumption that a single or a few, standalone challenges cannot always expose flaws of prepared or real-time deep fakes. Instead, this approach focuses on exploiting the vulnerabilities of offline and real time deepfakes by imposing a series of scalar difficulty, active and passive challenges to the impersonator. This method does not affect the experience and performance of a legitimate (human) user, but rather exposes flaws of deepfake tools used by impersonators and provides spoofing indicators with a satisfactory level of certainty.

The challenge set (i.e. the type and order of challenges selected) can be defined based on a series of factors:

- possible actions an applicant can perform in a particular live session;
- the environmental and ambient conditions;
- other security requirements.

Challenges contained in the database can be categorised as active and passive, with the following indicative examples of each category.

Active challenges:

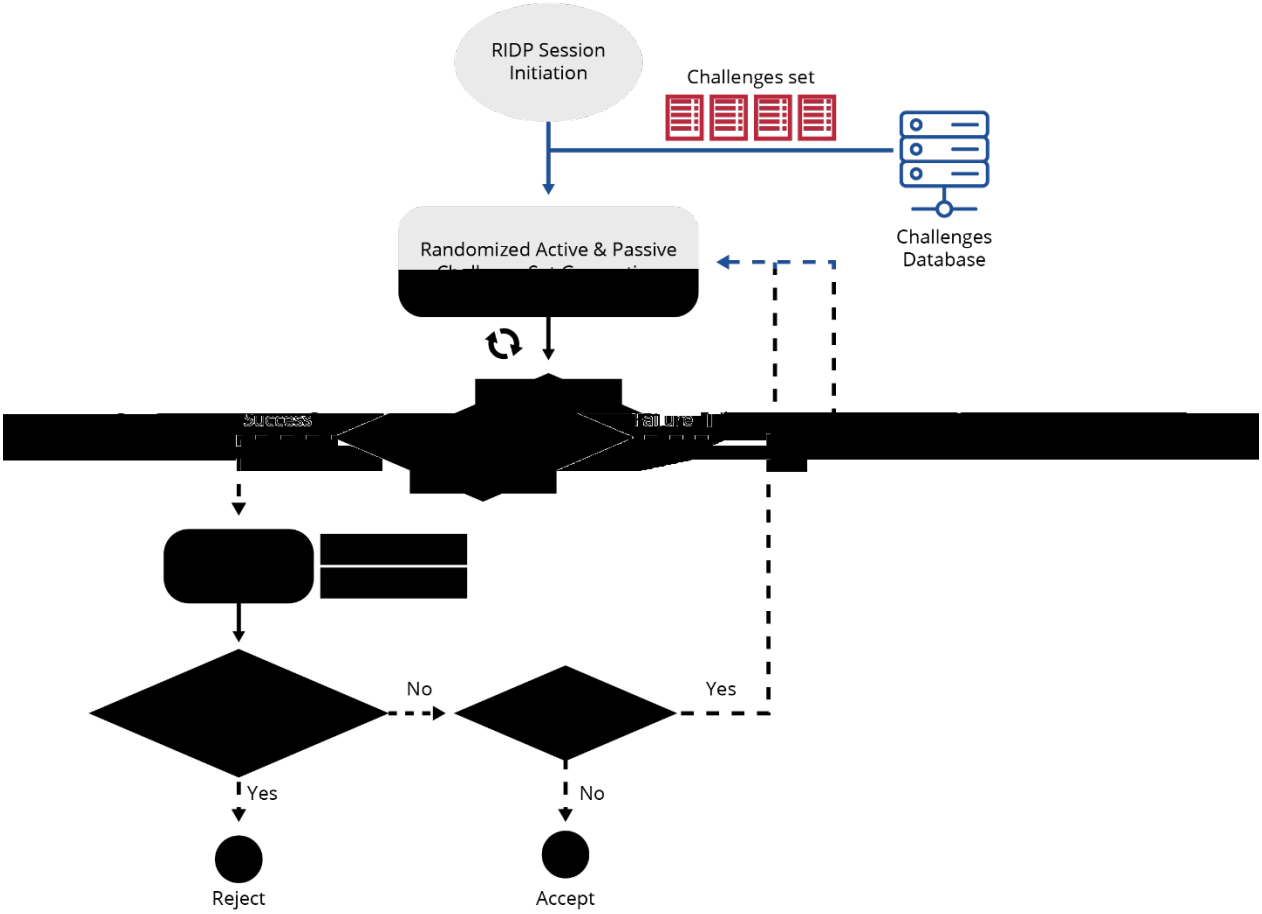
- occlusion, either using objects of the visible environment or synthetic overlays;
- intentional facial expressions to express a feeling or lip movements;
- intentional facial distortion (e.g. poking a cheek, revealing a part of the tongue);
- intentional alteration of the ambient lighting.

Passive challenges:

- involuntary facial micro expressions caused by specifically crafted visual stimuli;
- synthetic facial distortion;
- geometric synthetic distortions;
- synthetic alteration of ambient conditions (lighting, colour filters, light patterns).

Please note that especially for the passive challenges, it is assumed that a trusted camera or a mobile application is used, to ensure tampering protection of the client-side environment.

Figure 19: The high-level flow diagram of a combined active–passive challenge response mechanism



Multimodal biometric verification. By combining and verifying two or more biometric signals captured during the same identity proofing session (e.g. face and voice). Two points should be highlighted on this matter.

- Other types of biometrics (e.g. voice, fingerprints) are also susceptible to spoofing attacks (e.g. voice cloning). It should not be considered that a biometric different than the human face is not vulnerable or offers greater mitigation on its own.
- When a biometric system incorporates more than one biometric subsystem to perform multimodal verification, the false acceptance rate / false rejection rate performance of the weaker subsystem may compromise the security of the whole system. Thus, proper evaluation of the performance of subsystems should be made, depending on the joint decision mode of the overall system. For a detailed view on this topic, readers may consult the ISO/IEC TR 24722:2015 ([30]) technical report.

Human operator-based verification. A key observation made by ENISA during the data collection and analysis phases (surveys, face-to-face interviews, draft report validation) is related to the involvement of the human operator in the RIDP process. While there are fully automatic, AI-based RIDP solutions certified as a method offering a high level of assurance (LoA:High) as required by clause 24.1.d of eIDAS, the human operator is still considered a

([30]) https://www.iso.org/standard/64061.html.

crucial element of the RIDP process, especially for high-risk RIDSP cases. For most of the supervisory bodies interviewed, the current unattended automatic RIDP solutions are not sufficient, at least for critical, high-risk use cases; automatic methods and human involvement are considered complementary, with each able to detect different types of threats.

It is expected that more than 5 years of research are needed to validate the effectiveness of candidate countermeasures for digital injection attacks and to have a solid understanding of capabilities and limitations of the various approaches.

Thus, a risk analysis taking into consideration at least the following factors is needed:

- evolution of threat landscape (attack variants, methods, technology and tools);
- cost of human operators (continuous training, operation) versus cost of successful fraud.

A scheme providing gradual levels of assurance depending on the particular RIDP use-case and its criticality may help interested parties to better evaluate and drive the decision of choosing and combining the appropriate RIDP methods and technical solutions, spanning from unattended AI-based to hybrid.

Using trusted identification sources. According to some stakeholders, a very powerful tool would be the lookups to authoritative information sources (e.g. national database of identity documents) in order to verify if all the data presented during the RIDP process are correct. However, this is not available always to private organisations such as TSPs, because in many cases there is no relevant legislation to allow it, and such databases are only accessible by public administration. It was suggested that TSPs and RIDPs should be allowed to access such databases and registries in order to strengthen the RIDP process and create a harmonised approach across Europe.

Document status lookup. A first line of defence is to look up whether the document is marked as lost, stolen or expired by consulting national and international databases (e.g. SLTD, Public Register of Authentic Identity and Travel Documents Online). However, this measure cannot always prove effective since a document has to be already reported as stolen or lost in order to be properly detected during the lookups. Another obstacle is that lookups may not be permitted consistently among TSPs or that the lookups are performed against a limited dataset which corresponds only to the country of the company performing the lookup.

List of permitted document types. A list of the latest versions of permitted document types and their related acceptance criteria deployed at the national level could rule out documents that may provide assurance which is lower than a specific threshold, or documents with unknown security features or questionable validity. It is preferred that the specified documents are electronic and contain an NFC chip. This information should be shared among TSPs and RIDP providers around the EU.

Stricter requirements for the biometric NFC chip photo. Face photographs could be subject to stricter requirements, such as to reside in an encrypted form or to be stored as a set of multi-angle face photographs instead of a single shot, provided that adequate capacity is available in the NFC chip. An important initiative on this topic is the Coalition for Content Provenance and Authenticity ([31]), which aims to develop content provenance specifications for common asset types and formats to enable publishers, creators and consumers to trace the origin and evolution of a piece of media, including images, videos, audio and documents. Such a change would require further study and impact analyses on the specifications of passports and national identity cards with NFC chips.

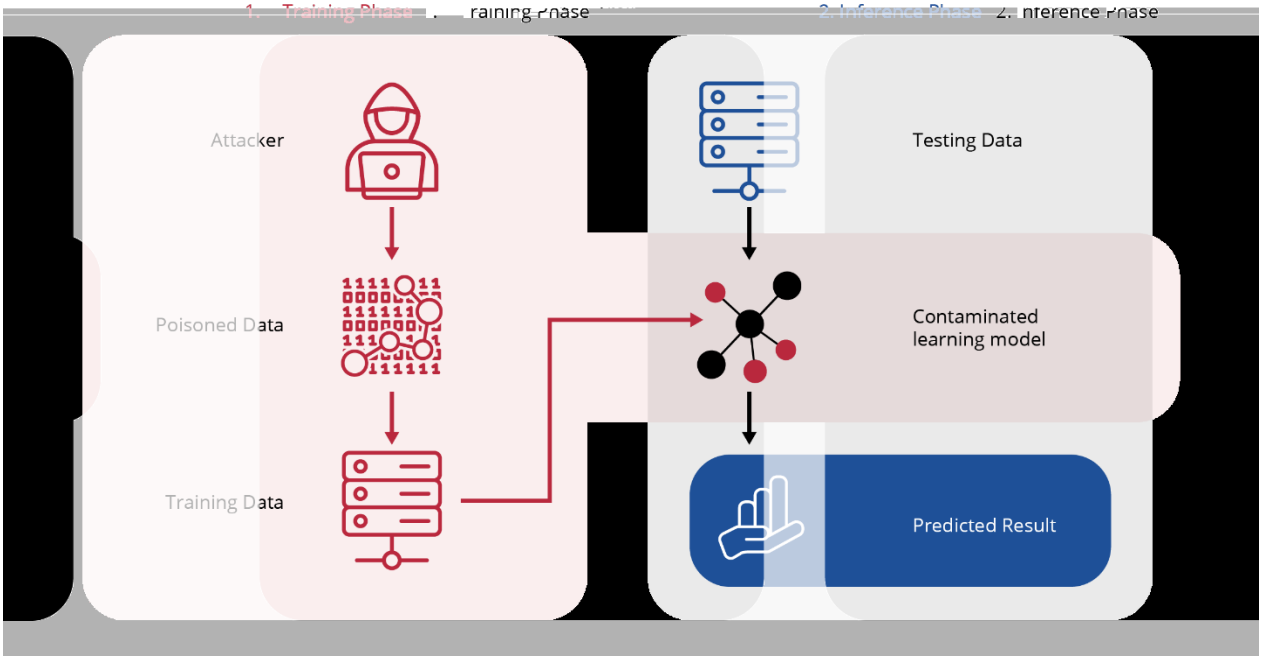
([31]) https://c2pa.org.

Threat monitoring. Another suggestion from the interviews was to focus on monitoring new attacks instead of focusing on the types of attacks. This will help authorities to adapt at the same rate of evolution and spot attacks as they are evolving, i.e. before they become successful.

As several RIDP vendors and TSPs are shifting towards automated AI and neural network techniques, it is possible that future attacks will focus on exploiting the weaknesses of neural networks. An example of such a new type of attack is the poisoning of training data, aiming to corrupt the training process by introducing false information to the training dataset of an machine learning (ML) system, which in turn can cause misclassification and wrong decision results.

Threat monitoring activities can be performed in a more coordinated and systematic way, in the form of a dedicated biometric security operations centre.

Figure 20: Diagram of a generalised ML poisoning attack



It is therefore possible that the potential new types of attacks will focus on circumventing AI systems instead of human operators. For this reason, the role of human involvement in RIDP processes may need to be expanded and/or revised.

4.6 ORGANISATIONAL CONTROLS

Apart from the technical and procedural good practices, wider organisational measures which can contribute to the strengthening of the security of RIDP are presented below.

Identity fraud-oriented risk assessment and treatment. Risk management is of paramount importance, not only in the context of RIDP but in the wider world of security and safety. Policies, procedures and implementation aspects of an RIDP service should be performed using a risk-based approach, taking into consideration not only generic cybersecurity and compliance risks, but also risks relating to identity theft and spoofing, such as behavioural attacks, identity document attacks and biometric face attacks (presentation, injection). Moreover, the risks should be revised regularly, taking into account the developments in the threat and technology

landscape, and should be treated according to a well-maintained and regularly updated risk treatment plan.

Alignment and adherence to recognised standards. Adherence to industry standards is the first obvious good practice falling under the wider organisation controls which can be followed by TSPs and RIDP vendors. This recommendation, however, poses a significant challenge since, in the domain of RIDP attacks, the standardisation landscape is not currently complete. Standardisation gaps have become obvious, with only one standardisation body having published a revised standard on presentation attacks (ISO/IEC 30107-3:2023). At the moment of writing this report, the domain of IAD standardisation remains incomplete (CEN TC 224 is working on a draft standard, see Chapter 2.4) while the threat landscape is rapidly evolving.

Use of tested software products and components. The use of biometric software products or components that have been thoroughly evaluated by competent laboratories provides detailed information regarding the assurance level, and the necessary visibility to implementors, auditors and supervisory bodies for RIDP solutions to be used in different contexts. Although accredited biometrics testing laboratories under ISO/IEC 30107 can perform the evaluation and issue a certificate for the PAD capability of a product, according to numerous stakeholders interviewed, the industry is still lacking a clear and widely adopted framework capable of determining the testing requirements, both for presentation and injection attacks.

Mandatory and recurring penetration tests, supplementary spoof bounty programmes. These are means to provide assurance through objective and high-expertise technical assessment operations, aiming to detect weaknesses, determine the robustness of controls and evaluate the security posture of biometric systems. Spoof bounty programmes (following the paradigm of bug bounty programmes) are public white hat security testing programmes designed to allow security researchers to identify and report vulnerabilities. Through this crowdsourcing practice, they can extend the assurance level in combination with formal testing and auditing, given that they are designed properly and reflect real use-cases of the RIDP product/service. The proper planning and combination of these practices can provide further assurance that testing is made as thorough and realistic as possible, without relying solely on the somewhat limited testing artefacts of a single testing laboratory.

5. CONCLUSIONS

Identity proofing, either in-person or remote, is a critical element of today's digital services and this is becoming more evident as the social, economic, regulatory and technological ecosystem is evolving. Developments such as those shown below confirm this assumption:

- the upcoming eIDAS 2 regulation;
- the various Member State legislation and certification schemes;
- the OECD's recommendations on the governance of digital identity;
- the rise of portable 'Know Your Customer' (KYC);
- the digital identity initiatives observed worldwide.

However, the remote nature of identity proofing is still not recognised equally at the national level in all Member States. At the same time, the way and the rate at which the technological landscape is evolving, both in offensive and defensive aspects, shows that identity proofing is the most-targeted element of digital identity.

According to this study's findings, a radical shift in the attack landscape has been observed, with digital injection attacks utilising deepfake technology considered the predominant type of attack, while presentation attacks are still used but to a far lesser extent. Factors that contribute to this surge of injection attacks are technological evolution, which allows easier production of fraudulent, synthetic identities, and the lower maturity of IAD methods in comparison to PAD methods, due to gaps in the available bibliography, scientific sources and standardisation on the topic of injection detection and mitigation. Based on attack insights collected during the preparation of this report, ENISA considers that a similar surge in sophisticated deepfake presentation attacks can occur anytime soon, resulting in a threat landscape with high complexity and sophistication, both in presentation and injection attacks.

Moreover, identity spoofing will soon not be the only attack objective. The scalability and level of automation of digital injection attacks, thanks to GANs, Crime-as-a-Service and the availability of deepfake generation tools, will introduce the denial of service (DoS) threat, not only against RIDP systems but also against interactive voice response systems / call centres. It is therefore evident that a shift in attack detection is required. Additionally, by 2030, deepfake technology will be considered the mainstream tool in the hands of criminal groups to launch targeted disinformation attacks. This insight has been further analysed in ENISA's report *Identifying Emerging Cyber Security Threats and Challenges for 2030* ([32]).

An important concern, identified in the early stages of the preparation of this report, is the lack of harmonisation among Member States regarding the functional and security requirements of the RIDP process, along with the evaluation methodology of such processes. Although RIDP is possible in most of the Member States, the exact circumstances under which it is permitted, and the applicable security and assurance requirements, are not uniform across the EU. While a few Member States have demonstrated an elevated maturity level regarding standardisation and regulation of the functional and security aspects of RIDP at the national level (e.g. France's PVID scheme, Germany's VDG and TR-03147, Spain's CCN-STIC 140), the majority of Member States have substantial or even ad hoc requirements with a variable level of detail, resulting in an inconsistent and unclear EU RIDP landscape. This may lead to market segmentation by national borders and unfair competition, due to the asymmetric nature of

([32]) https://www.enisa.europa.eu/publications/enisa-foresight-cybersecurity-threats-for-2030.

interpreting and applying cross-border trust services in the market, and may also have an impact on the operating costs of RIDP providers who operate in multiple Member States.

As also appointed by the report ([33]) of the Commission's Expert Group on Regulatory Obstacles to Financial Innovation, and eID & KYC Processes expert groups, national regulatory bodies of the various Member States impose different standards relating to technical aspects for digital identity verification.

The need to minimise polyphony and set a uniform baseline of requirements and permitted RIDP methods at the European level, taking into consideration the recent advancements of technology, becomes more evident than ever before. The balance between security and economic viability should be taken into consideration during the legislative processes, to avoid unnecessary compliance burdens. Additionally, as a note to standardisation bodies and policymakers, an effective effort to standardisation harmonisation in a rapidly evolving landscape would benefit from a looking-forward approach. Technology-neutral approaches that avoid describing specific technical requirements or solutions, but rather set the specific performance criteria, would lead to more flexible, adjustable and resilient standards published, with longer lifetimes and remaining up to date with the current state of the technological and threat landscape at every moment.

OPERATORS' EFFECTIVENESS BECOMES QUESTIONABLE in distinguishing between legitimate and synthetic identities.

Considering the rate and quality of fraudulent, synthetic identities generated by GANs, the difficulty for a human operator to distinguish them from legitimate ones will constantly increase, rendering the operator's effectiveness questionable. Multiple stakeholders shared the opinion of scanning the NFC chip of an authoritative ID document during the RIDP process. This method is considered to provide effective mitigation of impersonation attacks, since optical verification is not considered safe, due to the increasing sophistication of synthetic attacks. This particular measure is not universally applicable, since not all of the information stored inside the NFC chip can be legally accessed by private organisations (e.g. private TSPs, RIDP providers) across the EU. Another common opinion among interviewees was that efforts should be made towards the development of methods and tools capable of rapidly gathering threat intelligence, and subsequently building awareness for the various stakeholders of the ecosystem. This approach could enable visibility and effective mitigation planning of new threats. Moreover, the cost to keep human operators trained at all times will constantly increase, while the AI and deep learning algorithms capable of detecting the majority of synthetic identities will soon outperform the human operator. It is therefore suggested that human capabilities should focus on attack surface monitoring, threat hunting and identifying potential future threats and their attributes, along with sharing information across the ecosystem, to build more resilient and secure identity proofing systems and, finally, to break the constant arms race between offensive and defensive practices.

Complementary to the technical and information sharing-related controls and practices, the incorporation of lookups against trusted identity information sources (e.g. national registries) for the validation of the submitted information during the RIDP process was mentioned by various stakeholders. Currently, this cannot be applied in most cases, since such access to national identity information sources is not allowed for any private TSP or related organisation performing identity proofing. However, future changes in the EU regulatory landscape could set the basis for enabling controlled and secured access to national identity information sources, offering an additional weapon against identity spoofing. Such a practice is enabled by the digital verification services of the United Kingdom's Data Protection and Digital Information Bill ([34]).

Another issue expressed by some stakeholders relates to the binary nature of the trust services under eIDAS and the consequent effect on RIDP requirements. The fact that eIDAS defines only qualified and non-qualified trust services does not always fit well with the different levels of

([33]) https://digital-strategy.ec.europa.eu/en/library/reports-expert-group-eid-and-kyc-processes.
([34]) https://publications.parliament.uk/pa/bills/cbill/58-03/0143/220143.pdf.

assurance required in the various contexts. In other words, eIDAS established the electronic alternative of an ink signature to legally cover all cases, thus requiring the highest possible level of assurance. However, in many legitimate use cases, costs or ease of use are the main success factors, therefore a lower level of assurance is not only sufficient but rather preferred. This issue is considered an inefficiency of the eIDAS regulation that impacts, inter alia, the identity proofing process.

RIDP providers are compelled to support new methods of identity proofing, with increased accountability and legal duties, to support a variety of different use cases, some of which do not benefit from such a high level of assurance. This situation of 'one solution fits all' organically leads to compromises between security, ease of use and cost. For example, it may result in the adoption of RIDP solutions without diligent evaluation, and thus in TSPs with lower security posture in the context of RIDP.

An extension of the above issue was pointed out by some interviewed stakeholders, who suggested that a qualified signature for consumer credit reasons should not have the same weight as a qualified signature for an important medical operation. In the current implementation, it is believed that there is no proper way to express this differentiation, besides defining and setting a custom object identifier in the X.509 extended key usage field. In a risk-driven approach, a more granular model would be supported at the legal and technical levels to foster innovation and the wide adoption of RIDP solutions as a business enabler.

Overall, a more suitable and concrete approach would be to define a granular, technology-neutral contextual assurance scheme for qualified trust services, along with specific performance criteria for technical RIDP solutions, RIDP methods, providers and auditors, organised in different risk profiles per intended use cases. In that way, the assurance level will have a realistic meaning based on the context the service is realised, the accountability of the TSPs and RIDP vendors is more granularly distributed and the scheme is easily maintainable and adjustable to meet all current threats.

Important efforts towards updated standards on biometrics attacks have been made by ISO/IEC and the new version of 30107-3:2023 standard on testing and reporting on biometric PAD, along with the upcoming standard on biometric data IAD by CEN, setting the basis for future work on an international standard on biometric injection attacks.

While the constantly evolving attack landscape is becoming more and more complex, ENISA is continually working towards building awareness and producing risk-based analyses and reports, to support informed decision-making for the various stakeholders of the landscape and contribute to the development of countermeasures, helping RIDP to remain trustworthy and reliable in the years to come.

6. BIBLIOGRAPHY AND REFERENCES

6.1 BIBLIOGRAPHY

Agarwal, S., Farid, H., Fried, O. and Agrawala, M., 'Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches', *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, pp. 2814–2822, 2020.

Agarwal, A., Yadav, D., Kohli, N., Singh, R., Vatsa, M. and Noore A., 'Face Presentation Attack with Latex Masks in Multispectral Videos', *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, pp. 275–283, 2017.

Akhtar, Z., 'Deepfakes Generation and Detection: A Short Survey', *MDPI Journal of Imaging*, Vol. 9, No 1, 2023.

Baxevanakis, S., Kordopatis-Zilos, G., Galopoulos, P., Apostolidis, L., Levacher, K., Schlicht, I., Teyssou, D., Kompatsiaris, I. and Papadopoulos S., 'The MeVer DeepFake Detection Service: Lessons Learnt from Developing and Deploying in the Wild', *Proceedings of the 1st International Workshop on Multimedia AI against Disinformation*, Newark, pp. 59–68, 2022.

Busch, C., Nickel, C., Stein, C., Ramachandra, R., Raja, K., Wasnik, P., Stokkenes, M., Gomez-Barrero, M., Nautsch, A., Rathgeb, C., Scherhag, U. and Sousedik, C., *What Is A Presentation Attack and How Do We Detect It?*, Dan Panorama, Tel Aviv, 2018.

Carta, K., Barral, C., El Mrabet, N. and Mouille, S., 'Video injection attacks on remote digital identity verification solution using face recognition', *IMCIC, Proceedings of the 13th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2022)*, 2022.

Carta, K., Huynh, A., Mouille, S., El Mrabet, N., Barral, C. and Brangoulo, S., 'How video injection attacks can even challenge state-of-the-art Face Presentation Attack Detection Systems', *IMCIC, Proceedings of the 14th International Multi-Conference on Complexity, Informatics and Cybernetics (IMCIC 2023)*, pp. 105–112, 2023.

CEN/TC 224, *Biometric data injection attack detection*, working document, 2022-01

Chuming, Y., Wu, D. and Hong, K., 'Practical Deepfake Detection: Vulnerabilities in Global Contexts', *arXiv*, 2206.09842, 2022.

Dargaud, L., Ibsen, M., Tapia, J. and Busch, C., 'A Principal Component Analysis-Based Approach for Single Morphing Attack Detection', *2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, Waikoloa, pp. 683–692, 2023.

Diamant, N., Sandor, N. and Bronstein, A., 'Delta-GAN-Encoder: Encoding Semantic Changes for Explicit Image Editing, using Few Synthetic Samples', *Technion – Israel Institute of Technology*, 2021.

European Banking Authority, *Final Report – Guidelines on the use of Remote Customer Onboarding Solutions*, EBA/GL/2022/15, 22 November 2022.

European Union Agency for Law Enforcement Cooperation, 'Facing Reality? Law Enforcement and The Challenge of Deepfakes', observatory report from the Europol Innovation Lab, *Publications Office of the European Union*, Luxembourg, 2022.

Guarnera, L., Giudice, O., Guarnera, F., Ortis, A., Puglisi, G., Paratore, A., Bui, L., Fontani, M., Coccomini, D., Caldelli, R., Falchi, F., Gennaro, C., Messina, N., Amato, G., Perelli, G., Concas, S., Cuccu, C., Orrù, G., Marcialis, G. and Battiato S., 'The Face Deepfake Detection Challenge', *MDPI Journal of Imaging*, Vol. 8, No 22, 2022.

iProov Ltd., *Response to Request for Public Input*, PCAST Working Group on Generative AI, 2023.

ISO/IEC 30107-1:2023, *Biometric presentation attack detection – Part 1: Framework*, 2023.

ISO/IEC 30107-3:2023, *Biometric presentation attack detection – Part 3: Testing and reporting*, 2023.

Juefei-Xu, F., Wang, R. Huang, Y., Guo, Q., Ma, L. and Yang L., 'Countering Malicious Deep Fakes: Survey, Battleground, and Horizon', *International Journal of Computer Vision*, Vol. 130, pp. 1678–1734, 2022.

Korshunov, P., Chen, H., Garner, P. and Marcel, S., 'Vulnerability of Automatic Identity Recognition to Audio-Visual Deepfakes', Conference paper, *IEEE International Joint Conference on Biometrics*, *Idiap Research Institute*, September 2023.

Korshunov, P., Jain, A. and Marcel, S., 'Custom Attribution Loss for Improving Generalization and Interpretability Of Deepfake Detection', *ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, pp. 8972–8976, 2022.

Korshunov, P. and Marcel, S., 'Subjective and Objective Evaluation of Deepfake Videos', *ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toronto, pp. 2510-2514, 2021.

Li, C., Wang, L, Ji, S., Zhang, X., Xi, Z., Guo, S. and Wang T., 'Seeing is Living? Rethinking the Security of Facial Liveness Verification in the Deepfake Era', *arXiv*, 2202.10673, 2022.

Laas-Mikko, K., Kalvet, T., Derevski, R. and Tiits, M., 'Promises, Social, and Ethical Challenges with Biometrics in Remote Identity Onboarding', in Rathgeb, C., Tolosana, R., Vera-Rodriguez, R. and Busch, C. (eds), *Handbook of Digital Face Manipulation and Detection – From Deepfakes to Morphing Attacks*, Springer Cham, pp. 437–462, 2022.

Li, X., Komulainen, J., Zhao, G., Yuen, P. and Pietikainen, M., 'Generalized face anti-spoofing by detecting pulse from face videos', *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, pp. 4244–4249, 2016.

Lin, J., Dang, L., Rahouti, M., Xiong, K., *AI, Machine Learning and Deep Learning: A Security Perspective*, CRC Press, 2023.

Lyu, S., 'Deepfake Detection: Current Challenges and Next Steps', *arXiv*, 2003.09234, 2020.

Mao, M. and Yang, J., 'Exposing Deepfake with Pixel-wise AR and PPG Correlation from Faint Signals', *arXiv*, 2110.15561, 2021.

Mirsky, Y., 'DF-Captcha: A Deepfake Captcha for Preventing Fake Calls', *arXiv,* 2208.08524, 2022.

Mirsky, Y. and Lee, W., 'The Creation and Detection of Deepfakes: A Survey', *ACM Computing Surveys*, Vol. 54, No 1, pp. 1–41, 2020.

Mitra, A., Mohanty, S., Corcoran, P. and Kougianos, E., 'Detection of Deep-Morphed Deepfake Images to Make Robust Automatic Facial Recognition Systems', *2021 19th OITS International Conference on Information Technology (OCIT)*, Bhubaneswar, pp. 149–154, 2021.

Nanda, A., Shah, S., Jeong, J., Doss, R. and Webb, J., 'Towards Higher Levels of Assurance in Remote Identity Proofing', *IEEE Consumer Electronics Magazine*, Vol. 113, No 1, pp. 62–71, January 2023.

Negreiro, M. and Niestadt, M., 'Updating the European Digital Identity Framework', European Parliament, EU Legislation in Progress briefing, third edition, May 2023.

Pately, K., Hany, Hu, Jainy, A. and Ot, G., 'Live Face Video vs. Spoof Face Video: Use of Moiré Patterns to Detect Replay Video Attacks', *2015 International Conference on Biometrics (ICB)*, Phuket, pp. 98–105, 2015.

Qiu, W., 'A Survey on Poisoning Attacks Against Supervised Machine Learning', *arXiv*, 2202:02510, 2022.

Ramachandra, R. and Busch, C., 'Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey', ACM Computing Surveys, Vol. 50, No 1, pp. 1–37, 2017.

Rancha, S., Løvåsdal, G., Venkatesh, S., Raja, K., Ramachandra, R. and Busch, C., 'Analyzing Human Observer Ability in Morphing Attack Detection – Where Do We Stand?', *IEEE Transactions on Technology and Society*, Vol. 4, No 2, pp. 125–145, 2023.

Rathgeb, C., Tolosana, R., Vera-Rodriguez, R. and Busch, C., *Handbook of Digital Face Manipulation and Detection*, Springer Cham, 2022.

Scherhag, U., Rathgeb, C. and Busch, C., 'Face Morphing Attack Detection Methods', in Rathgeb, C., Tolosana, R., Vera-Rodriguez, R. and Busch, C. (eds), *Handbook of Digital Face Manipulation and Detection – Advances in Computer Vision and Pattern Recognition*, Springer Cham, 2022.

Scherhag, U., Rathgeb, C., Merkle, J., Breithaupt, R. and Busch, C., 'Face Recognition Systems Under Morphing Attacks: A Survey', *IEEE Access*, Vol. 7, pp. 23012–23026, 2019.

Seibold, C., Hilsmann, A. and Eisert, P., 'Feature Focus: Towards Explainable and Transparent Deep Face Morphing Attack Detectors', *MDPI Journal of Computers*, Vol. 10, No 9, 2021.

Sensity, *Deepfakes vs Biometric KYC Verification Report*, 2022, https://sensity.ai/blog/deepfake-detection/deepfakes-vs-kyc-biometric-verification/.

Sharma, D. and Selwal, A., 'A survey on face presentation attack detection mechanisms: hitherto and future perspectives', *Multimedia Systems*, Vol. 29, pp. 1527–1577, 2023.

Shoshan, A., Bhonker, N., Kviatkovsky, I. and Medioni, G., 'GAN-Control: Explicitly Controllable GANs', *arXiv*, 2021.02477, 2021.

Sun, Y., Zheng, J., Lyn, L., Zhao, H., Li, J., Tan, Y., Liu, X., Li, Y., 'The Same Name Is Not Always the Same: Correlating and Tracing Forgery Methods across Various Deepfake Datasets', *MDPI Journal of Electronics*, Vol. 12, No 11, 2023.

Taeb, M., and Chi, H., 'Comparison of Deepfake Detection Techniques through Deep Learning', *MDPI Journal of Cybersecurity and Privacy*, Vol. 2, No 1, pp. 89–106, 2022.

Veridas, 'Biometric Data & Biometric Systems: European & Spanish Legal Framework', 2022, https://veridas.com/docs/Veridas-Biometric-data-biometric-systems.pdf.

Wang, T., Liao, X., Chow, K. and Lin, X., 'Deepfake Detection: A Comprehensive Study from the

[13].Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A. and Ortega-Garcia, J., 'Deep Fakes and Beyond: A Survey of Face Manipulation and Fake Detection', *Information Fusion*, Vol. 64, pp. 131–148, 2020.
[14].FaceTec, *NIST FRVT-PAD Commentary*, 2022, https://facetec.com/FaceTec_NIST_FRVT-PAD_Comments.pdf.
[15].iProov, *Biometric Threat Intelligence Report*, 2023, https://www.iproov.com/reports/biometric-threat-intelligence.
[16].ISO/IEC 30107-1:2023, *Biometric presentation attack detection – Part 1: Framework*, 2023, p. 3.
[17].Gerstner, C. and Farid, H., 'Detecting Real-Time Deep-Fake Videos Using Active Illumination', *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, pp. 53–60, 2022.
[18].Ming, Z., Visani, M., Luqman, M. M. and Burie, J-C., 'A Survey on Anti-Spoofing Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices', *MDPI Journal of Imaging*, Vol. 6, No 12, 2020.
[19].Zhu, Y., Li, Q., Wang, J., Xu, C. and Sun, Z., 'One Shot Face Swapping on Megapixels', *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, pp. 4832–4842, 2022.
[20].Neekhara, P., Hussain, S., Zhang, X., Huang, K., McAuley, J. and Koushanfar, F., 'FaceSigns: Semi-Fragile Neural Watermarks for Media Authentication and Countering Deepfakes', *arXiv*, 2204.01960, 2022.

6.3 ENISA PUBLICATIONS

ID	Description
ENISA2022	ENISA, *Remote ID Proofing: Attacks & Countermeasures*, January 2022 https://www.enisa.europa.eu/publications/remote-identity-proofing-attacks-countermeasures
ENISA2021	ENISA, *Remote ID Proofing: Analysis of Methods to Carry Out Remote Identity Proofing Remotely*, March 2021 https://www.enisa.europa.eu/publications/enisa-report-remote-id-proofing

7. ANNEX A: GOOD PRACTICES OVERVIEW

ID	CONTROL DESCRIPTION	TYPE	FACE_PAD	FACE_IAD	ID_DOC
C-01	lighting conditions	ENV	X	X	X
C-02	minimum multimedia specifications (resolution, bitrate, microphone)	ENV	X	X	X
C-03	client-side architecture (dedicated application/web app, SDK/API)	ENV	X	X	X
C-04					X

8. ANNEX B: CHAPTER 3 EXAMPLES & FIGURES

Figure 21: Example of printed[1] and warped[2] photos

Figure 22: Example of a 2D printed and warped mask[2]

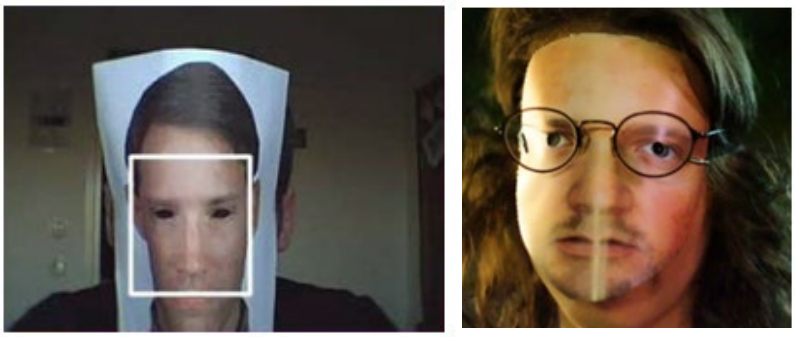
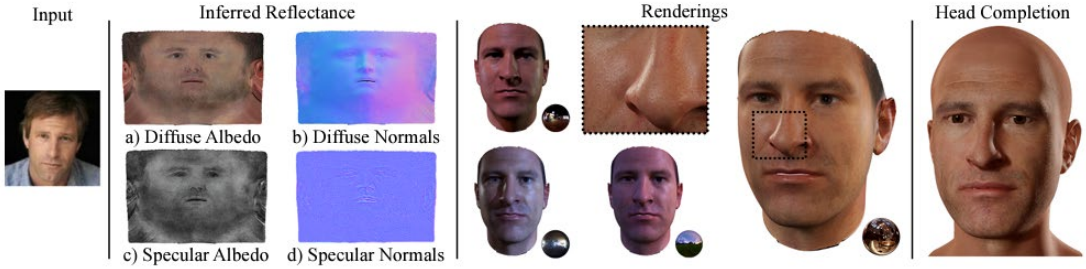


Figure 23: Example of a printed 3D layered mask[5]

Figure 24: Examples of resin 3D masks[2]

Figure 25: A 3D facial reconstruction of a 2D face photo[3]



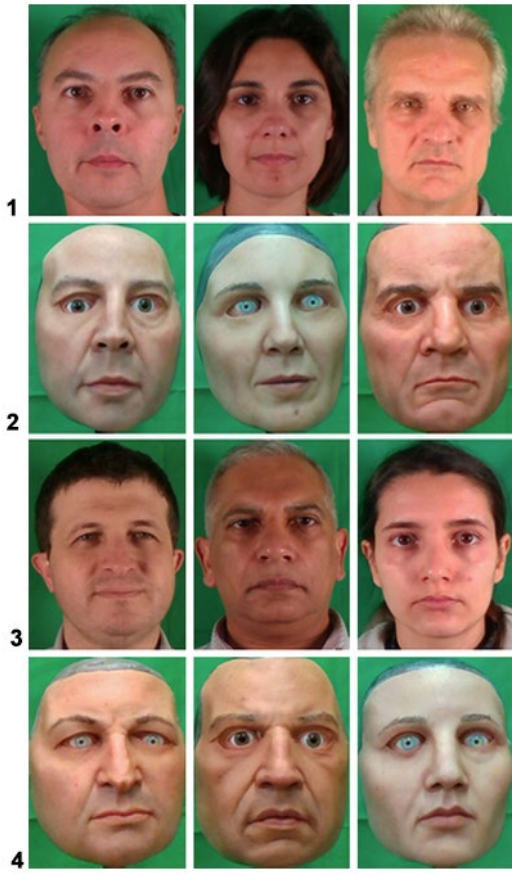


Figure 26: Silicone masks. Rows 1&3: bona fide presentations of the victimized subjects. Rows 2 & 4: masks corresponding to the subjects[4]

Figure 27: Examples of face video replay[1]

Figure 28: Example of a 3D face video render[5]

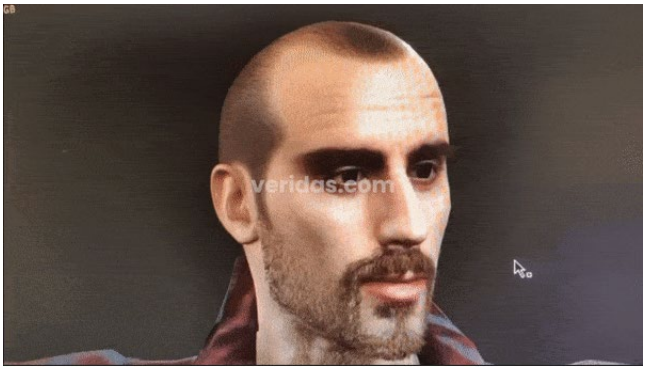


Figure 29: A face morphing algorithm[12]

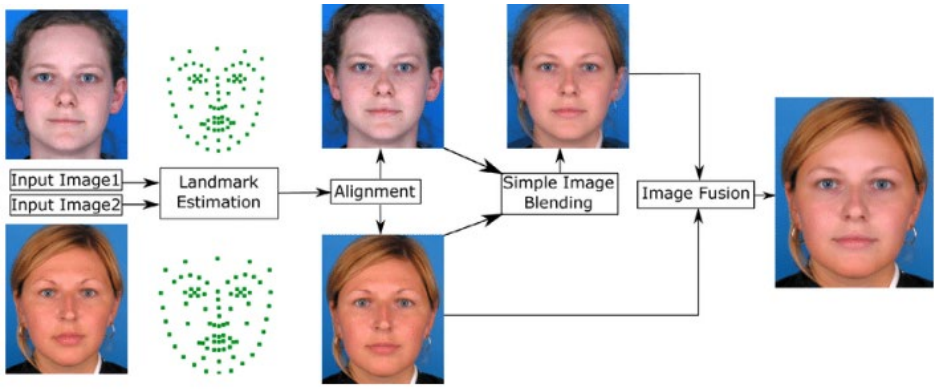
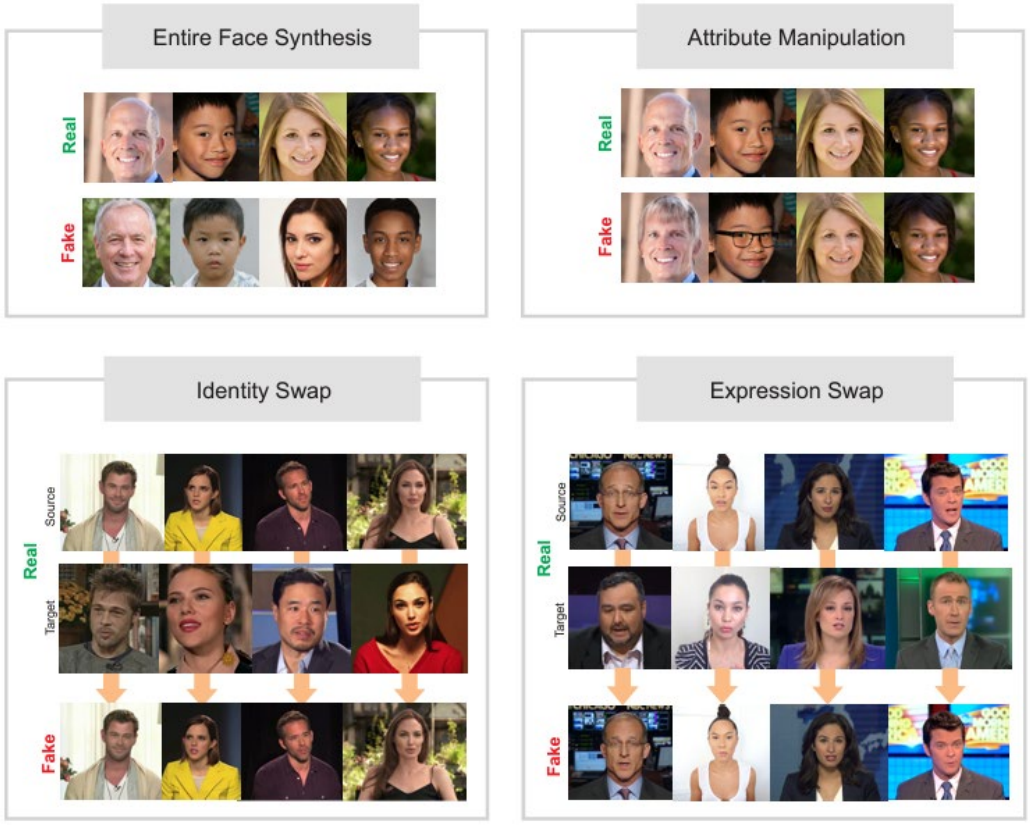


Figure 30: Real and fake examples of manipulation techniques[13]



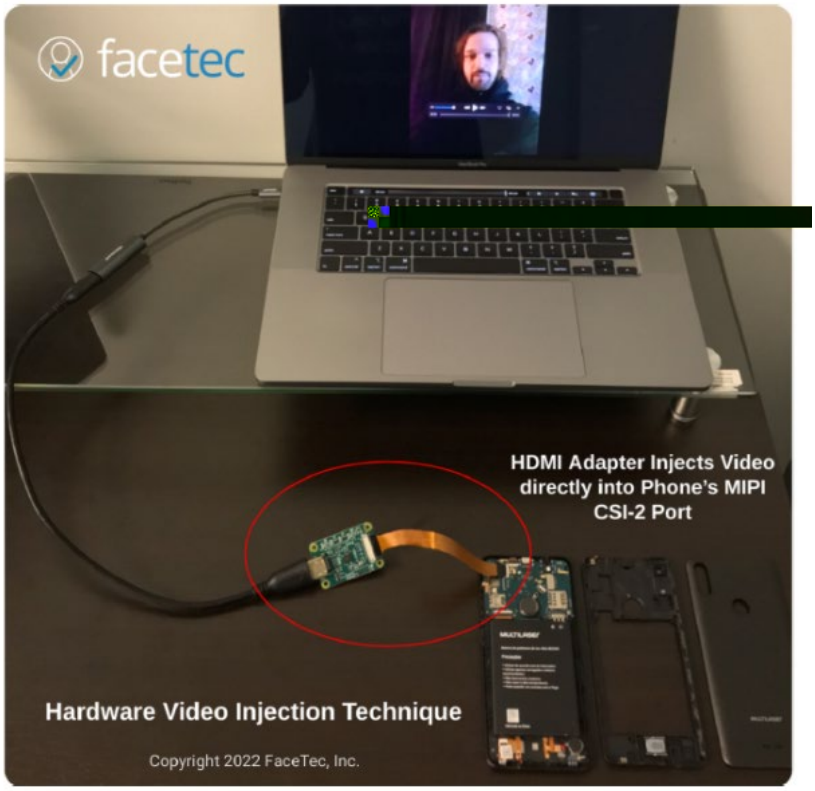


Figure 31: External video adapter injecting video from a computer to a mobile phone[14]

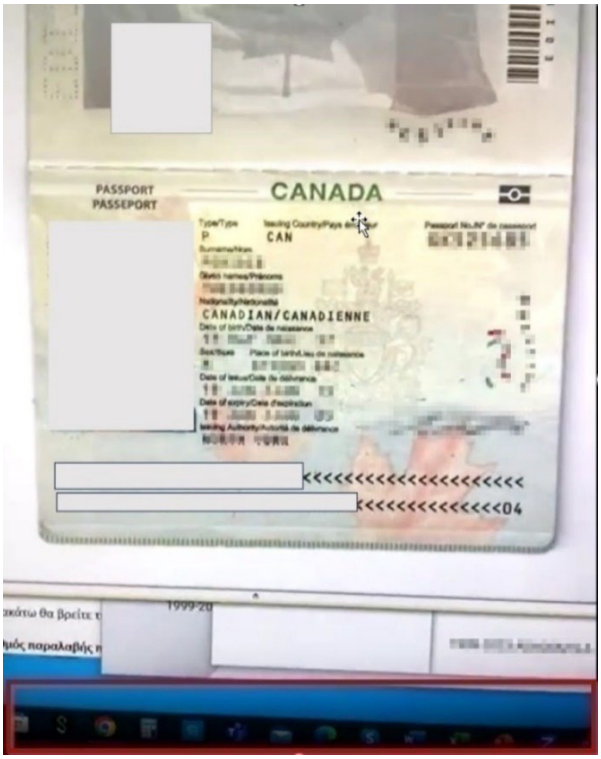
9. ANNEX C: REAL PRESENTATION ATTACK EXAMPLES

Below are two real examples from attempted presentation attacks, provided by a QTSP.

Example 1

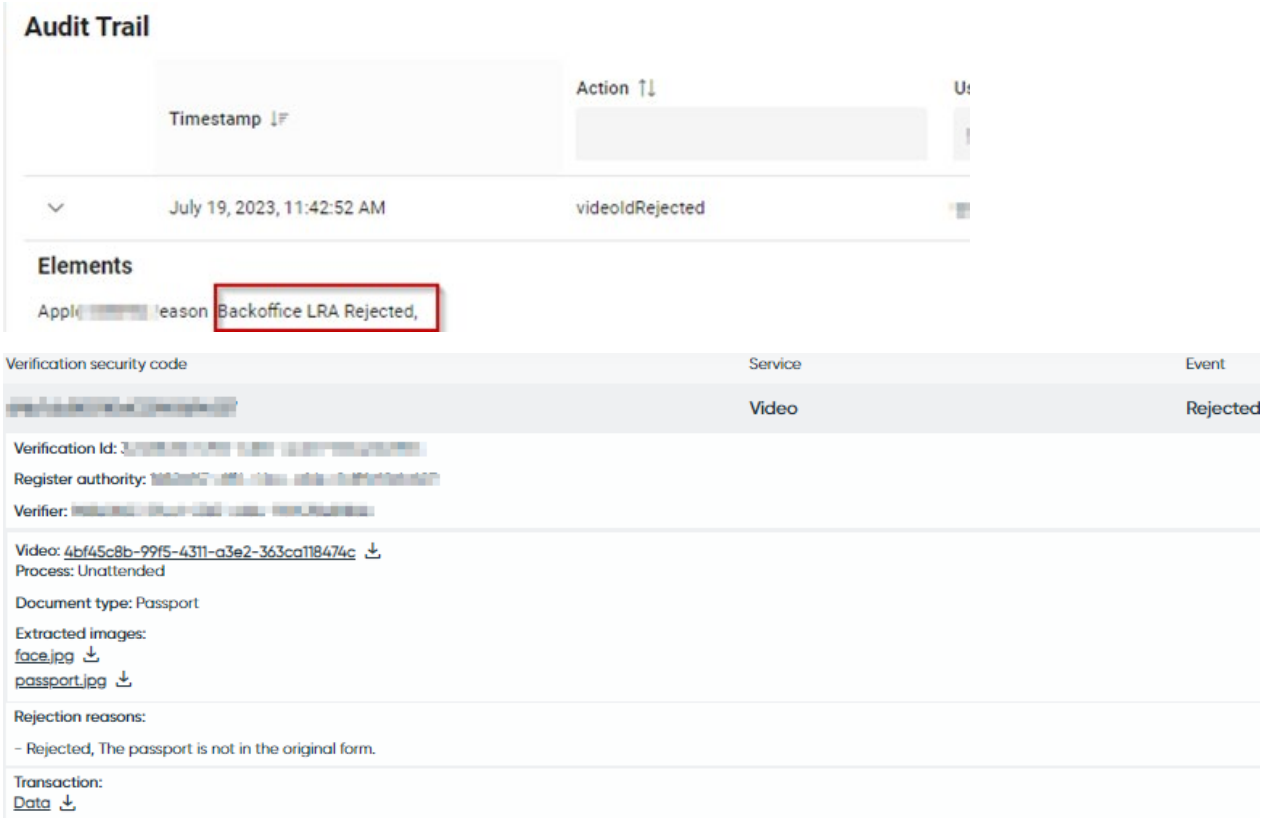
The first one, is an attempt from a Canadian person, who used a fraudulent passport. Although the document liveness checks did not classify the presented document as fraudulent, the human operator's review, led to the rejection of the document. The reason was the originality of the document could not be proven by the visual inspection performed by the operator, remotely.

Figure 32: Presentation Attack Fraudster's Fake Passport – Real Case 1



Through relevant logs recorded in the RIDP platform, the proofing actions and related information can be traced and audited.

Figure 33: Presentation Attack Audit Trail – Real Case 1



Example 2

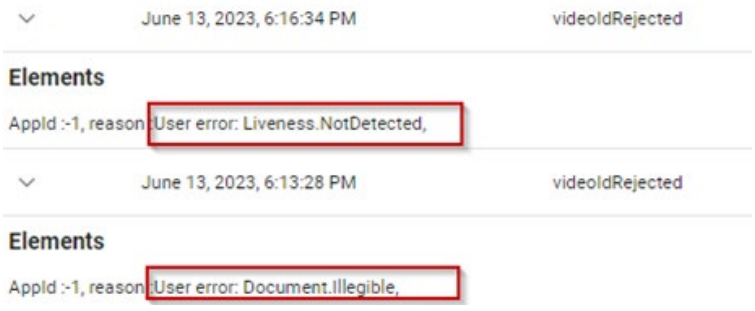
In this example, the document liveness checks performed by the RIDP platform were able to detect forged elements in the passport document.

Figure 34: Presentation Attack Fraudster's Fake Passport – Real Case 2

According to the logs of the platform, the following checks have failed:

- *Document.Illegible:* The document could not be scanned. Try from another device with
- better image quality
- *Liveness.NotDetected:* Live evidence could not be checked. Repeat the process following the instruction.

Figure 35: Presentation Attack Audit Trail – Real Case 2



TP-05-23-538-EN-N

ABOUT ENISA

The European Union Agency for Cybersecurity, ENISA, is the Union's agency dedicated to achieving a high common level of cybersecurity across Europe. Established in 2004 and strengthened by the EU Cybersecurity Act, the European Union Agency for Cybersecurity contributes to EU cyber policy, enhances the trustworthiness of ICT products, services and processes with cybersecurity certification schemes, cooperates with Member States and EU bodies, and helps Europe prepare for the cyber challenges of tomorrow. Through knowledge sharing, capacity building and awareness raising, the Agency works together with its key stakeholders to strengthen trust in the connected economy, to boost resilience of the Union's infrastructure, and, ultimately, to keep Europe's society and citizens digitally secure. More information about ENISA and its work can be found here: www.enisa.europa.eu.

ENISA
European Union Agency for Cybersecurity

Athens Office
Agamemnonos 14
Chalandri 15231, Attiki, Greece

Heraklion Office
95 Nikolaou Plastira
700 13 Vassilika Vouton, Heraklion, Greece

Brussels Office
Rue de la Loi 107
1049 Brussels, Belgium

enisa.europa.eu

ISBN 978-92-9204-661-3
doi: 10.2824/885606