EUROPEAN UNION AGENCY
FOR CYBERSECURITY

REMOTE IDENTITY PROOFING: ATTACKS & COUNTERMEASURES

JANUARY 2022

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

To contact the authors, please send an email to: team@enisa.europa.eu
For media enquiries about this paper, please use press@enisa.europa.eu.

EDITORS

Viktor Paggio, Evgenia Nikolouzou, Marnix Dekker

CONTRIBUTORS

Federica Magna, Piero De Simone, Jody Maganuco, Enrico De Waure, Fabio Galassi

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: Antonio Davoli, BacTech (Claude Barral), Bit4id (Paolo Campegiani), CEA-LETI (Jean-Francois Mainguet), Cleverbase (Bodine van Leeuwen), Digidentity (Marcel Wendt, Sander Remmerswaal) Electronic Identification (Carlos Paxarin, Carlos Asensio, Diego Burillo Aranda), FaceTec (Alberto Lima, Jay Meier, Kevin Alan Tussy), Galitt (Nathalie Launay), IDENTT, IDnow (Armin Bauer, Michal Kalinowski, Rayissa Armata), InnoValor (Bob Hulsebosch, Ines Duits), Innovatrics (Daniel Ferak), OneVisage (Christophe Remillet), SEALED (Sylvie Lacroix), SGM Consulting (Stephane Mouy), Thales DIS, Ubble (Caroline Goigoux, Juliette Delanoë, Nicolas Debernardi), Unissey (Sophie Finet), University of Cagliari (Fabio Roli), Yoti (Paco Garcia), Veriff (Janer Gorohhov). European Telecommunications Standards Institute (ETSI), ENISA Article 19 Expert Group, European Commission eIDAS Cooperation Network and various public authorities such as Danish Agency for Digitisation, French national Cybersecurity Agency ANSSI (Hugo Mania), German Bundesnetzagentur (Eva-Vanessa Ernst, Angelika Pfafferodt), Spanish Ministry of Economic Affairs and Digital Transformation, and others across the EU also contributed to this report.

LEGAL NOTICE

ISBN: 978-92-9204-549-4 DOI: 10.2824/183066 Catalogue number: TP-09-21-525-EN-N

EXECUTIVE SUMMARY

The past decade has seen rapid development in the field of information technology and a digital revolution that has provided unprecedented benefits to European society and its economy, by facilitating trade and the provision of services, creating new opportunities for businesses and boosting productivity and economic gain. Furthermore, the global pandemic underlined the importance of well-regulated remote identification processes and trusted digital identities that public and private sector organisations can rely on.

Remote identity proofing is a crucial element in creating trust for digital services. Remote identity proofing is the process where an online user proves he or she is the owner of a claimed identity. The proofing process is usually carried out over a webcam or a mobile device, where the users show their faces and produce their government issued documents – legal identity cards or passports. However, criminals are creative in devising tactics to circumvent these systems and a risk based approach is essential to stay ahead of the game.

Establishing a secure standardised environment for remote identity proofing could mitigate the risks but also bring benefits to organisations including better compliance, greater customer reach, competitive advantages, streamlined secure onboarding processes, whilst protecting users and their assets.

The present study analyses the collection and validation of evidence provided by the applicant to complete the verification of his or her identity. More specifically, **we focus on face presentation attacks that aim to fool the facial recognition systems.**

Potential threats to remote identity proofing methods were identified and listed, as well as the corresponding security controls, in the previous ENISA Report *Remote ID Proofing: Analysis of Methods to Carry Out Identity Proofing Remotely*[1] published in March 2021. Building on the previous report, this study will shed light on the details and practicalities of possible face presentation attacks against remote identity proofing methods, to better understand the existing threat landscape.

Through the analysis, which consisted of a review of the literature, a survey and interviews, **the following major face presentation attacks were identified:**

- **photo attack**
- **video of user replay attack**
- **3D mask attack**
- **deepfake attack.**

After different types of attacks were identified and classified, **applicable countermeasures were analysed and are presented in Chapter 4**.

Regarding different types of security controls, we focused on the overall security of the remote identity process, identifying the following control domains:

Environmental controls. Environmental checks refer to the hardware, software and network used by the user to carry out the process. An elementary control to consider is the verification of the video and audio quality level. This simple check can guarantee a better reception of information and consequently result in a more secure proof of identity. Another countermeasure that can be implemented concerns the execution of the process exclusively through a dedicated application. This allows, especially on smartphones, the implementation of different controls to check that the user's device is an actual physical object and that the camera feed is being

[1] 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

captured in real-time. Finally, the control of the metadata of the remote identity verification sessions, such as geolocation, IP, timestamps, VPN usage and others, allows fraud patterns and indicators to be identified.

Identity Document controls. The authenticity of identity documents is of fundamental importance within the remote identity proofing process. In the current state of technology, **the highest level of guarantee using a government-issued ID is represented by an electronic identity document equipped with an NFC chip.** The NFC chip contains the document data encrypted and digitally signed by the issuing state. Properly implemented video-based verification of modern identity documents featuring various security elements can also provide a reasonably high level of assurance and is useful when the users do not have access to an NFC ID or reader. For all types of documents, it is essential to verify that a document is not lost, stolen or expired by checking with national and international databases when access to such databases is available.

Presentation Attack Detection. The core of the automated processes is represented by the software used to perform Presentation Attack Detection (PAD). These software systems use artificial intelligence and machine learning to understand whether images were captured from a living human being and to do so they try to verify certain characteristics present in the images. **In terms of security, the video-based solutions commonly provide more data for analysis and therefore higher assurance of identity and that fraud is mitigated.**

The simplest PAD methods are based on expected face movements on command. Users are asked to perform random movements in order to verify that the video was not pre-recorded or is not being rendered in real-time by interactive deepfake puppet software. This approach is effective against unsophisticated attackers, particularly if the user is asked to perform fast movements as the software or underlying processing power will struggle to keep up. However, as widely available software and hardware performance improves quickly and approaches the frame rates of the average camera, other security controls need to be employed as well.

Other ways to check are based on involuntary human signals such as micro-movements and changes in the human face, eye movement, pupil dilation, micro-variation in the intensity of the skin colour given by the pulse of the blood and others. It is also possible to assess the three-dimensionality of the user via images acquired from different camera positions, stereoscopic cameras or through dedicated 3D depth sensors. Using multiple techniques allows the system to gather more information and make more accurate decisions.

Organisational controls. The technological component, even if central, is not the only one in the process and on the organisational side there are some controls that can be implemented. Within the organisational controls that can be implemented, the first is certainly to follow industry standards if available. To ensure good performance by operators and users, it is also important to design a straightforward and understandable remote identification process. To further harden the process, creating a spoof and camera bypass bounty programme to reward those who manage to evade controls can be effective. The controls should be well rooted in a risk-based approach and use a robust risk analysis methodology aligned with best practices to identify current threats and, above all, future and unknown ones.

Process controls. Controls on the execution of the process and its steps can be defined to make it more effective. It is important to perform periodic tests on systems and on the process. It is necessary to establish exactly what to ask the user during the remote identity proofing process, to save all the data relating to the single process for any future analysis (in compliance with the General Data Protection Regulation) and to identify exactly the actions to be taken in case of uncertainty about the result of the process, such as refusing identification or a request for intervention by an operator.

The future of attacks is a complex issue. We hope this report will benefit continuous structured risk analysis efforts in this field and **contribute to the development of countermeasures, helping remote proofing of identity to remain trustworthy and reliable** in the years to come.

TABLE OF CONTENTS

1. INTRODUCTION AND SCOPE

1.1 CONTEXT

Identity proofing is *the process by which a service provider collects and validates information about an applicant and verifies that collected and validated information actually belongs to the applicant*[2].

The classical way of performing identity proofing is for the applicant to provide evidence of their identity, such as presenting an identification document during a physical meeting with an operator of the service provider.

It is important to underline that remote identity proofing happens one step before identification and authentication, since it deals with the creation of identities under a set of very specific and restrictive circumstances, notably that the identity must be unequivocally bound to a physical person, which is not a requirement for most information systems.

In a physical world, the applicant and the operator need to be at the same place at the same time, a process which can be complicated, time consuming and, given the recent pandemic crisis, even dangerous for health-related reasons. Furthermore, proof of the validation of the evidence might not be properly recorded, and the operator might not be properly trained to perform correct verification of all the different kinds of acceptable evidence, or he or she could be psychologically manipulated, threatened, bribed, or otherwise convinced to improperly validate a false identity.

Methods to prove identity remotely provide the means to identify a person that eliminates the need for a physical presence, thus improving the user experience, reducing service provider costs, supporting the development of cross-border services, and avoiding unnecessary health risks. During the COVID-19 pandemic crisis, in fact, the possibility of identifying a person remotely became even more crucial.

Since remote identity proofing allows customers who are physically far away to access digital services, it opens a new level of business opportunities on the condition that identity proofing is performed in a trustworthy and secure way with a level of confidence equivalent to a face-to-face process.

Remote identity proofing can be used in a variety of contexts where trust in the identity of a natural or legal person is essential – be it in financial services, e-commerce, telecommunication services, the travel industry, human resources, matching platforms including delivery and ride-hailing services, public administrations, online gambling, and many other sectors.

EU Regulations on electronic identification and trust services – the eIDAS[3] – provide a common foundation for secure electronic transactions between citizens, businesses, and public authorities. Article 24 elaborates on identity proofing in the context of the issuance of qualified certificates, and paves the way for remote identity proofing by allowing *other identification methods... which provide equivalent assurance in terms of reliability to physical presence*.

[2] ETSI – *Electronic Signatures and Infrastructures (ESI); Survey of technologies and regulatory requirements for identity proofing for trust service, 2021-22:* *https://www.etsi.org/deliver/etsi_tr/119400_119499/119460/01.01.01_60/tr_119460v010101p.pdf*
[3] *Regulation (EU) No 910/2014 on electronic identification and trust services for electronic transactions in the internal market:* *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG*
[1] Please use footnotes for providing additional or explanatory information and/or relevant links. References should be listed in a dedicated section. Use only the function References/Insert Footnote

The European Commission is currently evaluating eIDAS and ran an open consultation from 24 July to 2 October 2020[4]. The aim of the consultation was to collect feedback on drivers and barriers to the development and uptake of trust services and eID in Europe. The study also considered the impact of the options for delivering an EU Digital Identity.

However, remote identity proofing is used far beyond the context of trust services and a means for electronic identity as defined in the eIDAS or the financial sector as covered by the 5th EU Anti-Money laundering directive (AMLD5[5]). Thus, the scope of this report is not limited to the eIDAS ecosystem but encompasses technologies in current use and related security recommendations regardless of their legislative setting.

Proofing identity remotely is a crucial element in enabling trust for digital services. In the previous ENISA Report *Remote ID Proofing: Analysis of Methods to Carry Out Identity Proofing Remotely*, potential threats to remote identity proofing mechanisms were identified and listed as well as their corresponding security controls. Building on the previous report, and in line with ENISA's mission to serve as a centre of network and information security expertise, this report will shed light on details and practicalities of possible attack vectors against remote identity proofing methods.

This report will also validate the security controls proposed in the previous report *Remote ID Proofing: Analysis of Methods to Carry Out Remote Identity Proofing Remotely* and provide further recommendations on how to mitigate identified threats, to create a more secure cyber environment and boost the uptake of remote identity proofing and related services.

1.2 TARGET AUDIENCE

The report is aimed primarily at the following stakeholders:

- EU companies and other public or academic organisations that run or prepare to launch their own remote identity proofing solution;
- EU companies and other public or academic organisations, including national governments and various public bodies, who are considering implementation of a remote identity proofing solution for their clients, citizens, employees, students, and other stakeholders, or those who have already implemented it and want to make their solution more secure;
- more specifically in the eIDAS ecosystem:
    - Trust Service Providers and Identity Providers that might use this report to harden cybersecurity of their own remote identity proofing solutions,
    - Conformity Assessment Bodies and Supervisory Bodies that evaluate (or supervise the evaluation of) remote identity proofing solutions or Trust Services using those solutions
- security researchers, academia, and the wider security community.

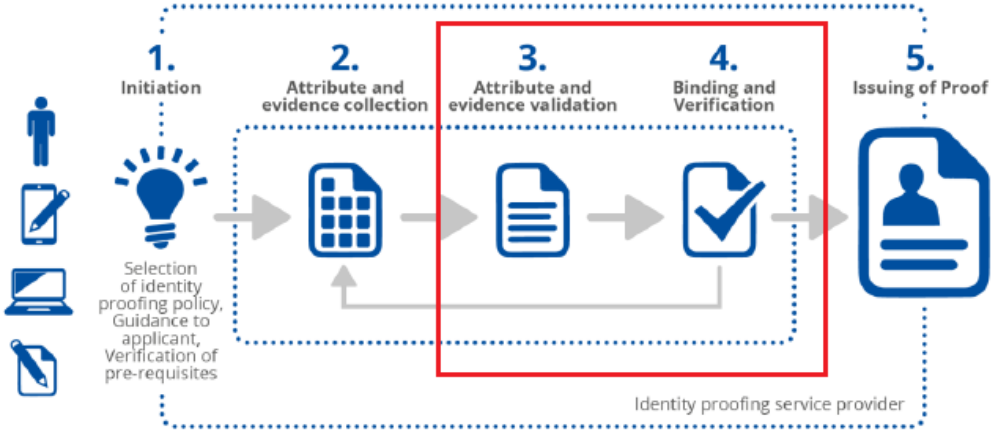
1.3 STRUCTURE AND SCOPE

The goal of this report is to describe attack techniques against remote identity proofing mechanisms, with a focus on attacks on face presentation, to validate the security controls proposed in the previous ENISA report and provide further practical countermeasures to mitigate such attacks.

Let's consider for a moment the generalised five (5) steps diagram for an identity proofing process described in the previous ENISA report:

[4] *https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eIDAS-Node+version+2.0*
[5] *Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU (Text with EEA relevance): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018L0843*

Figure 1: Identity proofing process



To go through this process, an applicant requires:

- definitive proof (usually, photo or video evidence of his or her face) that the applicant is physically present with the device and ID
- a government issued ID attesting his or her identity, which is to be validated by the service provider
- a high confidence match between the liveness-proven photo or video evidence and the face shown on the ID document or in the NFC Chip

Once the match is confirmed, the service provider can 'bind the document including all the data it contains to the applicant.

Consequently, an attacker whose goal is to fool the system by impersonating someone else (or creating a synthetic identity) has the choice of spoofing:

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

Since ID forgery is an already well-known and widely studied problem that has found a technical solution through the use of public key infrastructure (PKI) to authenticate data stored electronically in microprocessor chips, such as in e-passports, this report will focus mostly on attacks that target the binding phase by spoofing the face of the applicant, either through physical or digital means such as silicon masks and deepfakes.

The focus of the report is therefore related to the very first moment of onboarding when the user introduces himself to the system, their physical characteristics are compared with their ID and the verified data are consequently tied together and saved for future use. Subsequent interactions are out of the scope of this report, though when liveness-proven biometric data has been verified to belong to a specific legal identity, that biometric data can be used again and again in the future to re-verify that user without the need to re-process the ID documents again.

Please notice that face comparison is not the only way for a service provider to bind an identification document to an applicant. Other methods could theoretically be used for remote identity proofing, such as fingerprints, retina scan, blood vessels in hands, voice recognition, etc., but these would require:

- for this kind of biometric data to be stored in the identity document of the applicant,

- for the applicant to have a device able to read this information both from his or her document, such as an NFC reader, and from his or her physical features, for the service provider to compare them,
- strong detection of liveness to ensure the biometric data is captured in real time directly from a live human being.

Fingerprints are more and more found in identity documents based, for example, on recent EU regulations on strengthening the security of identity cards[6], but the data will be only accessible to authorised persons and only when the document is required to be produced by law. Hence, since the use of fingerprints, retina scan and other methods mentioned above is not applied in most real-world scenarios, they have not been considered in the scope of this report.

Other aspects that have not been analysed in this report, as they would conflict with its technical and practical nature. These include:

- human-related internal threat scenarios regarding operators of a remote identity proof 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;
- generic cyberattacks aimed at underlying technologies (user endpoint stations, servers, tampering with data in transit given improper encryption, etc.) or human factor (generic social engineering attacks etc.), except where a close and direct impact on remote identity proofing methods is specifically observed and explained;
- the *hows* and *whys* of the illustrated attack scenarios including, for example, how attackers may obtain stolen or forged documents, how they may obtain high-quality silicone masks, how to inject video into a camera feed, where to download deepfake creation or virtual camera software, etc.;
- viable, but implausible scenarios such as attacks performed by doppelgangers, plastic surgery to impersonate someone else, etc.;
- privacy and data protection issues related to personal data, biometric data processing etc.

Concerning the methodology implemented, data collection and the aggregation of the results, a mixed approach was adopted. This mix implies above all complementarity between qualitative and quantitative evidence gathered through desk research, interviews, surveys and workshops. Surveys and interviews were conducted with stakeholders across four categories outlined during the initial desk research and questions were customised for each category of stakeholder.

Questions for stakeholders from technology or identity providers were focussed on gathering practical information based on the experiences of their company or organisation. Meanwhile questions posed to stakeholders from categories such as academia and national certification bodies investigated the opinions of experts regarding new threat scenarios and possible countermeasures in more detail.

To report evidence and the opinion of stakeholders on the topic of attack scenarios on remote identity proofing and validate what was found during the desk research, all information was aggregated which also ensured the confidentiality of the information obtained. More details on the methodology used are illustrated in Annex A, and in Annexes B and C the results of the survey and workshop are illustrated.

This report follows a straightforward and logical structure, which begins with an analysis of the existing remote identity proofing methods (Chapter 2) to identify applicable attack methods

[6] *Regulation (EU) 2019/1157 of the European Parliament and of the Council of 20 June 2019 on strengthening the security of identity cards of Union citizens and of residence documents issued to Union citizens and their family members exercising their right of free movement*, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019R1157*

aimed at fooling them (Chapter 3) and ultimately to provide practical and actionable countermeasures (Chapter 4).

Figure 2: Report logical structure

Section 2 - Methods

Regarding remote identity proofing methods, this report builds on the comprehensive analysis already included in ENISA's previous report on *Remote ID Proofing: Analysis of Methods to Carry Out Remote Identity Proofing Remotely*, while focusing on the specific steps and methods that were deemed relevant to achieve its own goals and objectives.

This report does not include analysis of remote identity proofing methods that are based on electronic identification or digital certificates. These methods, in fact, rely on a digital identity that has already been proofed by a third party and don't extend beyond the device to the physical human user of the device. Attacks affecting such methods rely on generic cyberattacks targeting underlying technologies, such as third-party compromise, stolen credentials, or stolen or expired certificates, instead of attacks designed to deceive remote identity proofing systems, and thus were not considered interesting nor within the scope for this report. The method that involves the use of an electronic identity document and related PIN was also not considered, as this also requires a prior proof of identity and communication between the systems involved.

Section 3 - Attacks

Regarding attack methods, the analysis performed in this report encompasses those currently used in the wild, those whose feasibility has been validated by security researchers, and those still in conceptual stage, but with a realistic probability of being introduced in the future. The study focuses on face presentation attacks.

Following desk research, four main types were analysed:

- **photo attack**
- **video of user replay attack**
- **3D mask attack**
- **deepfake attack**

Section 4 - Countermeasures

Finally, the report presents countermeasures against the attack methods analysed, be they of a technical or organisational nature, based on a risk-based approach. Countermeasures are described in practical terms and actionable intelligence is presented, whenever possible, to support the stakeholders to whom this report is addressed.

Supplementary material

Additional material related to the methodology used and the main results of interviews, surveys and workshops can be found in Annexes A, B and C. Finally, Annex D shows a map of the methods, attacks and countermeasures.

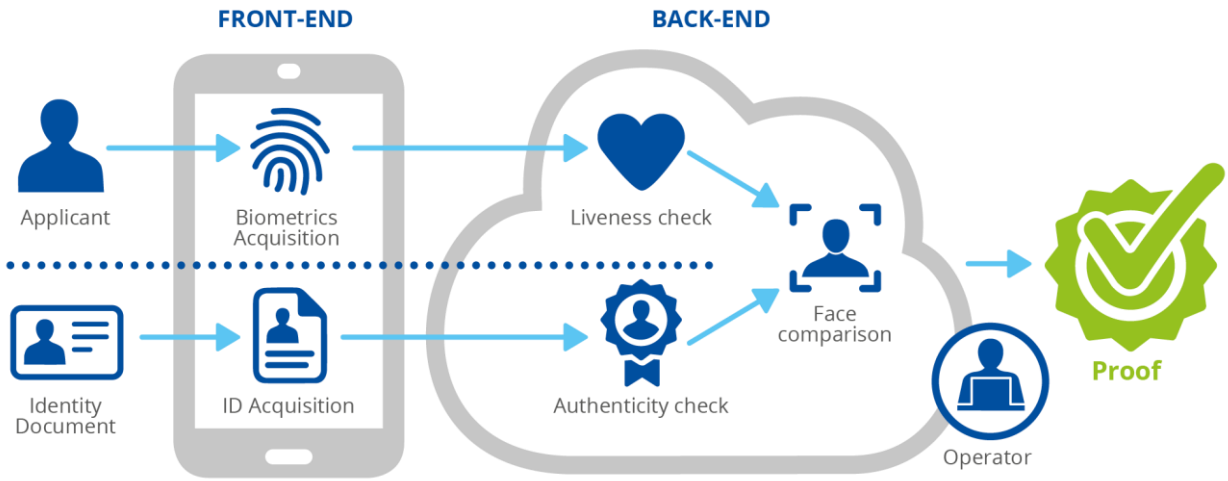
2. REMOTE IDENTITY PROOFING METHODS

Building on the comprehensive analysis already included in ENISA's report *Remote ID Proofing: Analysis of Methods to Carry Out Identity Proofing Remotely*, this chapter will discuss various remote identity proofing methods to lay out the foundations required to understand the different kinds of attacks.

The analysis will focus on practical, real-world implementations of the remote identity proofing methods, collected during numerous interviews with technology and service providers, in order to discern different possible implementations and the pros and cons associated with each one.

Most remote identity proofing methods can be modelled through the following diagram:

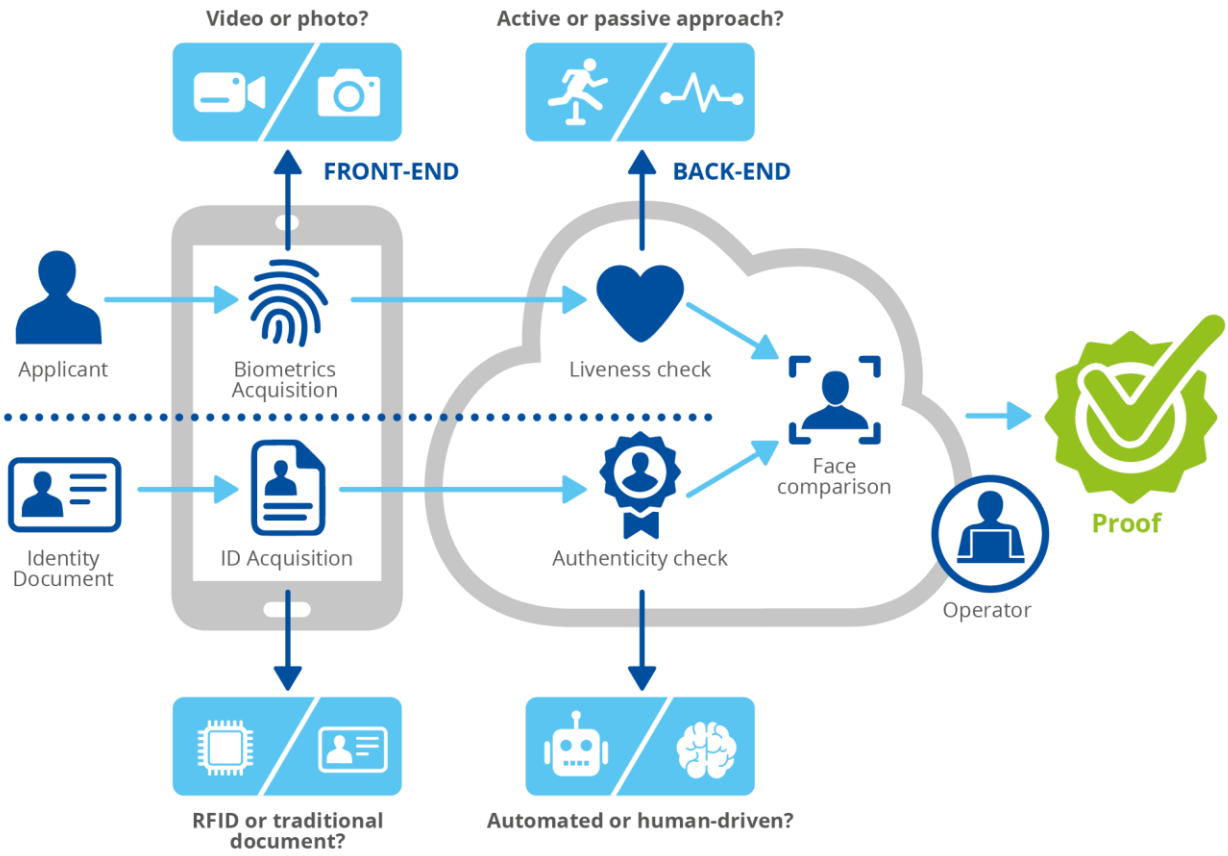
Figure 3: Generic proofing method diagram



Many differences were observed during the interviews as to how each service provider approaches and implements each step of the process. Some of these differences may seem subtle at first but they have a huge impact on the security and confidence of the identity proofing process.

The most significant and frequently observed differences have been noted in the following diagram.

Figure 4: Differences observed in proofing methods



The following subsections provide a brief description for each phase shown in the diagram, discussing, whenever applicable, the different approaches that have been observed during the interviews to provide a practical view of real-world implementations of remote identity proofing methods while analysing the pros and cons of each approach.

Please note that the goal of this analysis is not to determine which implementation is generally the best, since different use cases may favour different approaches and there is no optimal solution in the Paretian sense, i.e. there is no implementation where no evaluation criterion (e.g. security, cost, accessibility, useability, etc.) can be made better off without making at least one other criterion worse off.

2.1 BIOMETRICS ACQUISITION

The process always begins with an applicant enrolling to a remote identity proofing (RIDP) service, usually to access and use digital services from the RIDP provider itself or a third party. The enrolment process could also be used to obtain a digital identity.

The applicant will interact with the service provider through a front-end, which in most cases is a mobile app from the service provider that the applicant needs to download onto his or her smartphone or a web application accessed through a web browser on a computer.

Through the functionalities available from this front-end, the applicant is required to provide evidence of:

- their facial features, usually through the acquisition of video or one or multiple photos;

- a valid, government issued identity document from which will be created the face reference either by using the photo stored in the chip (for an electronic document) or using the photo as printed on the document (when no chip or other encrypted secure media is available).

This evidence is usually transmitted through an encrypted channel and processed in the RIDP application back-end which is operated by the service provider.

When it comes to biometrics acquisition, the first significant difference in approach between the remote identity proofing methods observed is regarding the method of choice for this task.

Some service providers carry out this task by processing a **full video comprised of 100 of more frames** of the applicant's face, others capture **one or multiple frames** (usually between 2 and 4).

Each approach has its own merits. **In terms of security, the video-based solutions provide more data for analysis and therefore higher assurance as to identity and the mitigation of fraud.** Proponents of the photo approach believe that the benefits of the video approach in terms of a false acceptance rate are not significant enough to justify the additional downsides in user scenarios they cater to.

One of the most obvious differences between these two approaches is in the quantity of data that must be transferred between the front-end and the back end of the application, which can range from a few kilobytes required to transmit a frame to the many megabytes required for a full video. The difference in data transmitted can range between 2 and 3 orders of magnitude and may be significant for applicants in countries where high speed connections are not widely available or expensive, thus limiting the pool of potential customers for the service. In short, lower data transfers mean higher inclusiveness for the potential user base, as well as lower environmental costs caused by data transmission and storage.

The additional data to be acquired and processed may also play a role, considering the number of requests per day, in terms of computing resources required from the service provider, leading to higher operational costs.

However, some technology providers process the applicant's biometric data on his or her device instead of on their servers, and only send up the applicant's specific data (3D image, textures etc.) along with video session metadata. In that case, the high resolution video of the user's face never leaves the device and the payload sent to the server is significantly smaller. In this way, the video-based approach can also be leveraged in environments with bad connectivity or where data prices are steep.

Ultimately, no size fits all – it is up to each organisation to pick the technology that is matched to its own needs, user scenarios and risk appetite as well as taking specific market conditions into consideration, including low bandwidth or steep prices for data.

Nowadays in Europe, harmonised certification schema or benchmarks are yet to be developed. However, some member states such as France have gone a step beyond proposing, at national level, a *Remote identity verification providers – Requirements Framework*. Regarding the requirement for evidence as part of the identity process, the standard specifies that this must be a video and sets a minimum resolution and frame rate[7].

[7] The minimum resolution, after compression, shall not be less than 720p: 1280 × 720 at 25 frames per second[7]. See: ANSSI, *Référentiel d'exigences applicables aux prestataires de vérification d'identité à distance,* March 2021,
[1] Please use footnotes for providing additional or explanatory information and/or relevant links. References should be listed in a dedicated section. Use only the function References/Insert Footnote

Video as evidence is also suggested as a requirement in an ETSI standard[8] that suggests recording a video sequence with ID movement as a possible mechanism to verify that it is real and authentic.

There are more interesting initiatives, such as the European TReSPAsS-ETN[9] consortium of universities and industrial entities, which focus on delivering improved detection technologies for presentation attacks.

2.2 BIOMETRICS LIVENESS CHECK

This phase encompasses all the controls that are performed on the applicant to obtain reasonable assurance that the subject in front of the camera is a real and live person and not, for example, a deepfake puppet, a mask, or a reprojection of a pre-recorded video or photo.

The controls performed during this phase can be incredibly exhaustive and complex according to the kind of technology and process automation used by the service provider. For example, some of the service providers that were interviewed in the making of this document reported that their liveness checks include more than 90 AI-powered controls on the video evidence provided by the applicant.

Attacks on these controls are called presentation attacks and will be further explored later in the chapter on attacks.

Two important concepts that must be introduced regarding these controls are **Attack Presentation Classification Error Rate** (APCER) and **Bona fide Presentation Classification Error Rate** (BPCER):

- **APCER** – measures the percent of attack presentations incorrectly classified as bona fide presentations, i.e. the proportion of imposters who should be rejected, but nevertheless succeeded and passed the liveness check;
- **BPCER** – measures the percent of bona fide presentations incorrectly classified as attack presentations, i.e. the proportion of legitimate applicants labelled as spoofs.

Since no control system can ever be 100% secure, the objective of each RIDP system is to achieve balance between these two parameters that is adequate for its purpose.

Reducing the APCER to the lowest possible level, in fact, will likely result in a sharp rise of the BPCER. In other words, the more secure the control system, the less convenient it will be, as users are falsely rejected by the system. The same also applies the other way round.

The maximum acceptable level of APCER for a specific RIDP system largely depends on the use case for which the system has been deployed, since specific use cases can be way more security-sensitive than others (e.g. identification for the public administration v e-commerce).

Details regarding these controls are rarely shared by service providers, due to business decisions but also the fact that they cannot prove and measure negative events, and it is extremely difficult to test all possible variations of spoof artefacts and attack vectors. AIs are not transparent today, and it is hard to measure their performance as few independent, unbiased

https://www.ssi.gouv.fr/entreprise/actualite/publication-du-referentiel-dexigences-applicables-aux-prestataires-de-verification-didentite-a-distance-pvid/

[8] ETSI – *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for trust service components providing identity proofing of trust service subjects,* TS 119 461, BIN-8.4.2-01.
https://www.etsi.org/deliver/etsi_ts/119400_119499/119461/01.01.01_60/ts_119461v010101p.pdf

[9] TReSPAsS-ETN is a consortium of seven universities (beneficiaries), supported by seven industrial entities (partner organisations) located in France, Germany, Netherlands, Switzerland, Spain, and Belgium, *https://www.trespass-etn.eu/consortium.html*

datasets and testing grounds for liveness are available. A high-level overview and some recommendations are provided in the Countermeasures section.

Nevertheless, two approaches have been observed.

- **Active liveness detection**: the user must align their face with the camera, typically in an oval or a reticle, and is then asked to perform specific actions in response to challenges such as nodding, blinking, smiling, reading random words, following an object moving on the screen, getting their face closer or farther from the camera, turning their head left or right on command, or waving a hand in front of their face. The response must match what is challenged.

- **Passive liveness detection:** also requires the user to align their face with the camera, typically in an oval or a reticle, but the AI relies exclusively on verification techniques that require no user action, such as inconsistencies in resolution resulting from deepfake manipulation, flashing lights on the subject to analyse reflections on the applicant's face and surroundings, analysing microvariations in the intensity of skin colour due to pulses of blood and other involuntary human signals present in the analysed video frames or images.

The two approaches therefore differ in the active or passive participation of the user but not in the evaluation of the result. In fact, both the acquisition of evidence and the decision as to whether the user is a real and living person **can be performed by an artificial intelligence, human operator, or both.** Both approaches also require verification that the device is physical hardware, not an emulator, and that the camera is indeed capturing the video feed in real time.

Many of the service providers interviewed believe that the active liveness approach further increases the difficulty and cost of fraud, since it is harder for fraudsters to work around these kinds of controls with techniques such as photo attacks, video replay attacks and deepfakes. At the same time, the active detection of liveness may have a negative impact on new customer acquisition, as users can get frustrated by the challenge-response nature of the process, which in turn leads to abandonment rates which may be significant for the service provider.

Passive liveness detection, on the other hand, provides the most transparent and seamless customer experience for the applicant, but may be not as effective in deterring injected photo attacks, video reprojection attacks and deepfakes.

Real-world implementation scenarios may favour one solution over the other and often involve a mixed approach that combines the strengths of both these techniques and on-device security checks of the camera feed.

2.3 IDENTITY DOCUMENT ACQUISITION

A significant differentiating factor in real-world implementations of RIDP methods regards the types of identification documents that are considered acceptable for the applicant to provide during the enrolment phase. This, in turn, has huge implications for the technologies that must be put in place in order to acquire those documents and on the control system required to obtain reasonable assurance on the authenticity of those documents.

A significant difference exists between traditional, **paper-based documents** and **electronic documents**.

Electronic documents, the most common being e-passports, use an embedded electronic microprocessor chip which contains biometric information that can be used to authenticate the identity of the holder. They mostly use contactless smart card technology, including a microprocessor chip (computer chip) and antenna (for both power to the chip and communication) embedded in the document, even though contact chip documents requiring special smart card readers are also in use.

The critical information is printed on the document, repeated on the machine-readable lines (MRZ) and stored in the chip. public key infrastructure (PKI) is used to authenticate the data stored electronically in the chip, making it borderline impossible to forge with current technologies when all security mechanisms are fully and correctly implemented.

Figure 5: Electronic passport

Furthermore, the communication channel between the chip and the reader is protected by encryption: before data can be read from a chip, the reader needs to provide a key which is derived from the machine-readable zone, which prevents data being stolen without visual access to the passport (e.g. scanning it covertly while still inside user's pocket).

Some identity documents, such as the German and Italian documents, add a security measure for the use of some online features, requiring the use of a PIN delivered with the document when it is issued.

Working with electronic documents requires the deployment of optical character recognition (OCR) technology from the service provider to read the MRZ needed to access the chip, and a device able to read the microprocessor chip from the applicant, like an NFC reader. Also, even though contactless chips gain prominence, many countries still have contact chip eID cards to be used with a dedicated card reader.

The consensus on electronic documents is that they are pretty much tamper proof and provide the highest level of assurance on the identity of the holder. At the same time, an RIDP method relying only on electronic documents would be very limited in reach and accessibility, since:

- worldwide, the diffusion of electronic documents is still severely limited compared to traditional ones,
- NFC readers are not widely available on older and low tier smartphones

and thus, is usually reserved only for the most security-sensitive use cases.

Traditional, paper based documents, on the other hand, make the service widely accessible even if they provide a multitude of challenges as most of them were not designed with remote

identification in mind. The level of assurance they can bear is proportionate to the number and quality of their security features.

For enabling high level verification of authenticity of any document without a chip or other encrypted secure storage, the verification should require a video of both the front side and back side of the documents while asking the documents to be tilted to check variable optical features. The video should be of high quality to enable a reference photo to be extracted of sufficient quality for facial comparison and to verify any counterfeiting of personal data or the presentation of any forged documents, either by physical or digital means. The verification of such a video requires both specialist automatic means (e.g. to check for covert visible features such as micro text or other high security printing) and human verification by an expert operator.

One of the main issues with traditional documents is the vast number of different types of documents that exist. Multiplying the number of countries, issuing authorities, valid documents (e.g., ID card, driving license, etc.) and different versions of those documents together can easily lead to thousands of acceptable document types. Some of the service providers interviewed have reported building an internal database of acceptable document types that exceeds 10.000 entries. In this regard, public registers of authentic identity and travel documents like European PRADO prove to be very useful[10].

Some of these documents, often the recent ones, have physical security features, such as holograms, but a significant percentage of them have ineffective security features or no security features whatsoever. Also higher quality fake documents can have holograms, anti-tamper lines, perforation and other security features. Therefore, implementing detailed checks that can provide full assurance on the authenticity of every single one of those documents remains impossible for any service provider.

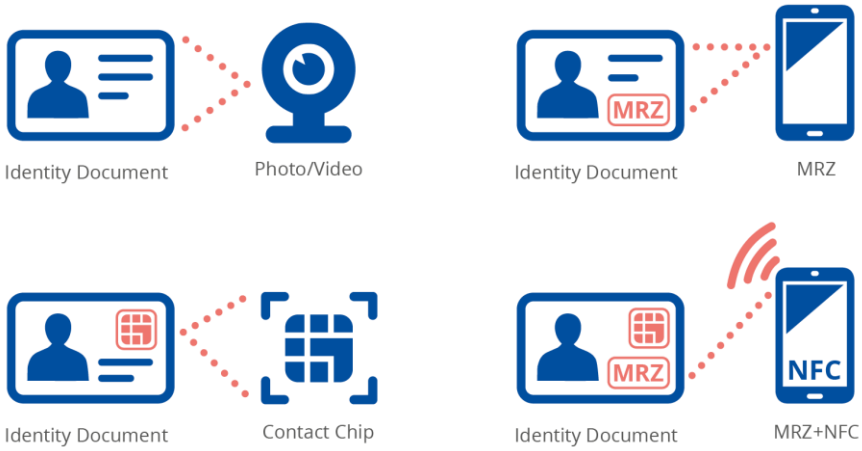
The main types of identity documents with the relative characteristics that allow the acquisition of their data can be summarised as follows:

- traditional paper based
- MRZ based
- contact chip based
- NFC based

Often, the line between types is blurred – for example a traditional paper based document can have an MRZ that is checked in parallel to the security features.

[10] Council of the European Union – *Public Register of Authentic Identity and Travel Documents Online (PRADO), https://www.consilium.europa.eu/prado/en/prado-start-page.html#*

Figure 6: Types of IDs and their characteristics



In the end, deciding which documents shall be acceptable for an RIDP implementation is not a clear-cut decision, but again a delicate balancing act between the accessibility of the system and the assurance level it enables.

2.4 IDENTITY DOCUMENT AUTHENTICITY CHECK

This phase includes the controls which are performed on the applicant's identity document, to obtain reasonable assurance that it is a real, authentic and valid document issued by a government and is not a forgery or spoof.

The controls performed during this phase obviously change according to the kind of identity document presented.

Figure 7: eID's data authenticity diagram

In case of an electronic document, as anticipated in the 'ID acquisition' paragraph, public key infrastructure (PKI) is used to authenticate the personal data and the user photo stored electronically in the chip. Also, the authenticity of the chip is verified against cloning.

If PKI is implemented correctly then with current technologies it is not possible to violate the underlying crypto suites. However, the quantum computers of the future will be able to do so – hence the call for fast implementation of quantum-resistant algorithms. Also, PKI mechanisms do not provide full proof that the correct living human is holding the device at the given moment.

In case of traditional documents, evidence is usually captured through the camera of the applicant's device and several authenticity checks are performed on it, including data validation checks through OCR technology and visual, pixel-level analysis aimed at spotting inconsistencies and indicators of forgery or the swap of a face photo.

Liveness check of the video feed can also be active or passive. Analogously to a face liveness check, the applicant can be asked move his or her identity document (active) or only passive controls are employed (passive). The decision on the user's physical presence with the device can be performed by AI, a human operator or both (hybrid approach).

Again, details regarding these controls are not shared by service providers and are beyond the scope of this report, but a high-level overview and some recommendations are provided in the Countermeasures section.

2.5 FACE COMPARISON

This phase is performed only if both previous verifications, i.e. of user liveness and ID document verification, have been successful. Its purpose is to obtain reasonable assurance that the person identified by the ID is the same person in front of the camera, thus effectively binding the person with the document and greenlighting him or her to obtain his or her digital identity.

One last differentiating factor between real-world implementations of RIDP methods is the **level of automation** and, consequently, **human involvement** regarding this step and the whole RIDP process.

While completely manual and software-assisted processes have been considered within the scope for this report, all the service providers interviewed have moved away from those models and implemented fully automated models or hybrid solutions where human operators intervene systematically or on-demand.

In this phase, two concepts are used to measure system performance, **False Acceptance Rate** (FAR) and **False Rejection Rate** (FRR):

- **FAR** – measures the percent of invalid inputs that are incorrectly accepted as valid
- **FRR** – measures the percent of valid inputs that are incorrectly rejected as invalid.

Rapid advancements in Deep Neural Networks have not only made the automated controls faster and cheaper but, according to most service providers, also more effective than their human counterparts. **Some service providers claim a False Acceptance Rate for the comparison of real user faces to their photo ID that is almost a full order of magnitude lower in favour of AI** (0.3% vs 2-3% according to one of the interviewees).

In the automated model, human intervention is required only in limited cases, i.e. where the percentage of certainty with which AI is recognising the applicant is below a certain pre-established threshold, and the operator is asked to confirm the correctness of the recognition.

That does not mean that humans do not play a crucial role anymore in RIDP processes, as humans are still largely responsible, for example, for the development and training of those AIs. Furthermore, as multiple interviewees pointed out, AI might be better than humans at a very specific task, such as comparing pictures of faces, but is not very good outside of that specific task, such as analysing the context in which those pictures were taken, for example, spotting contextual elements, which are apparent to humans, and which might indicate that the applicant is being threatened, tricked or coerced into the RIDP process if, say, the applicant is unconscious or somebody is holding a pistol to his or her head. This may change as service providers continue to accrue larger datasets on these scenarios but for now human operators are still very much needed.

Some stakeholders in the industry believe that humans will have a *bigger* role in RIDP processes in the coming years for a different reason. As the service providers are shifting towards automated techniques, many believe that future attacks will shift focus on techniques aimed at exploiting the weaknesses of the neural networks, including, for example, adversarial attacks.

Adversarial attacks in the remote identity proofing space may leverage adversarial contamination of training data, i.e. contaminate the datasets used for training remote identity machine learning models and then use specially crafted, deceptive inputs which cause misclassification and wrong decisions. Or, through systematically probing the face comparison model, attackers could circumvent the security controls directly without access to training data. As the focus of attack techniques might shift from fooling humans to fooling AIs, the role of humans in RIDP *processes may see an increase* in importance.

There is no silver bullet in terms of remote identity proofing methods. Which techniques should be adopted and which controls should be implemented depend vastly on the use-cases for which the system is designed and on the level of assurance that it is expected to provide, following a risk-based approach.

3. ATTACKS

The remote identity proofing methods illustrated are based on comparing the face photo on the identity document or contained in an electronic document chip with the applicant's liveness-proven face.

The face is not the only type of biometric data. The fingerprint, the physical conformation of the hand, the physical conformation of the iris or retina, the vocal timbre, etc are also types of biometric data. These data would all be valid for the unique recognition of an individual but at present, as already discussed, the face image is used due to the technical characteristics of mobile devices and desktop computers that are used to provide evidence.

The attack methods related to the identity document are not the subject of this report, but it is important to illustrate the feasible scenarios as the identity document is essential evidence for the remote identity proofing methods in current use.

The attack methods related to the identity document are:

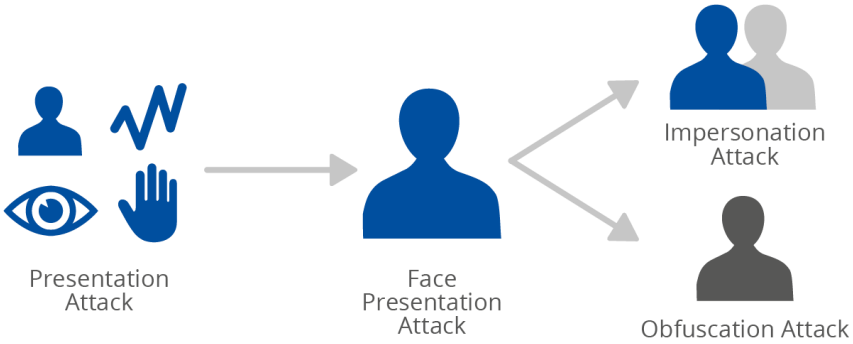
- authentic identity document with one or more modified parts (stolen, expired, etc.);
- complete reproduction of the identity document of a real identity;
- complete production of an identity document for a fictional identity;
- complete reproduction of an identity document of a partially real, partially fictional identity, for example with the date of birth changed;
- fantasy identity document created from scratch without reference to an existing type of document.

The forgery of the identity document can be physical (changes made to the physical document) or digital (changes made to the photo or video of the document). Regarding digital counterfeiting, online digital tools to make fake identity documents are available[11], plus there are software leveraging technologies such as deepfake that make the identification of changes on a human-based visual basis practically impossible and call for controls of another nature.

The attack type that tries to deceive biometric recognition is called **Presentation Attack** and in ISO / IEC 30107 it is defined as *presentation to the biometric capture subsystem with the goal of interfering with the operation of the biometric system*. When the attack relates to the biometric data of the face this is called **Face Presentation Attack**.

[11] Verif Tools – *Passport Online Generator,* available online

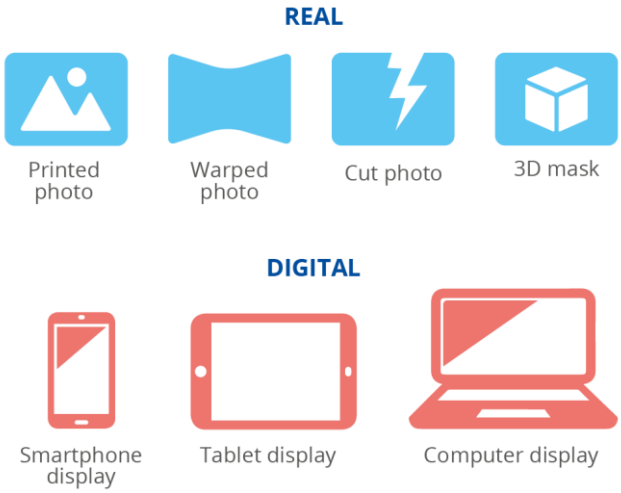
Figure 8: Face presentation objectives diagram



A presentation attack is performed with one of the following two objectives: impersonation or obfuscation. We speak of impersonation when the attacker's goal is to use an identity other than his or her own (real or fictitious) and obfuscation when the attacker's goal is to avoid being recognised and circumvent the system. **The object of this report is impersonation**. Obfuscation cannot occur in the remote identity proofing methods being considered in this report, as the process cannot be completed.

With impersonation attacks we can describe a wide range of different threats. Each attack has a **PAI (Presentation attack instrument)** to exploit the vulnerability, which can be either real or digital.

Figure 9: Presentation attack instruments

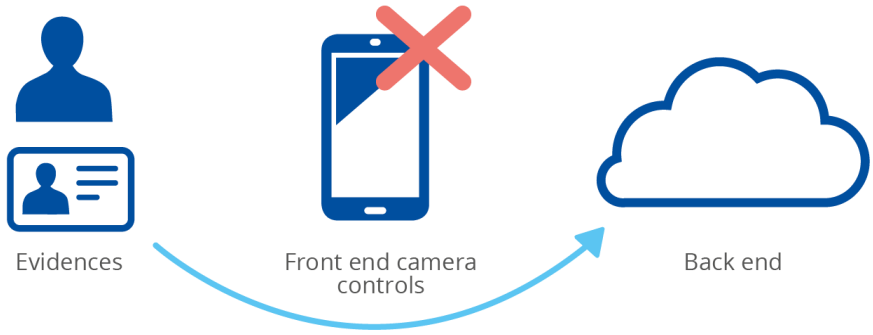


Digital presentation attack instruments are basically displays, that can be used to reproduce real or modified digital photos and videos. Using robust algorithms it had been previously possible to identify the use of screens in PAD attacks, but now with 4-8k HDR screens, which are higher resolution than the cameras being used to capture the images and do not show pixels and other artefacts of manipulation, this type of attack has become extremely viable. So, providers need to implement detection methods such as the 3-dimensionality detection of face liveness, detection of screen glare and others.

Physical presentation attack instruments, on the other hand, are those that can literally be touched by the human attacker. This category contains the different methods used to simulate a face, such as silicone masks or printed masks and faces. The attacker can directly hold these objects in front of his or her face, or wear them, and try to fool the system.

While the instruments shown are popular for attacking attempts, video injection is the most promising attack method, according to market players. It is not a presentation attack instrument, as the video will be injected and not physically presented to the camera. It is hence a way to circumvent the cam of the device used to perform the remote identity proofing.

Figure 10: Video Injection



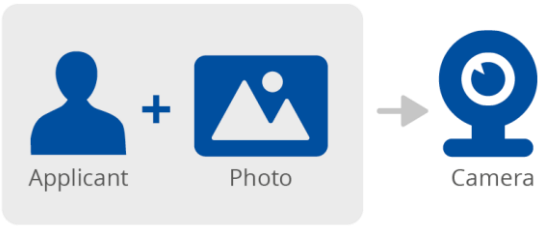
The actions an attacker must take to perform an attack are the same regardless of the type of attack:

- choose the identity to be used,
- find data on the identity to be used,
- find or create an identity document consistent with the identity to be used,
- create the identity in relation to the type of attack and the PAI to be used,
- present the identity to the system.

3.1 PHOTO ATTACK

The attack is based on the presentation as facial evidence of an image of a face printed or displayed via a device screen.

Figure 10: Photo attack diagram



This kind of attack can be used against systems which use photo or video as evidence. Basically, the attacker will put the photo in front of the camera to let it acquire the information needed. This is an extremely simple method that most liveness AI service providers can detect consistently by focusing on the differences in texture and brightness from human skin, the lack of dynamic shadows, lack of depth and other spoof artefacts.

Figure 12: Samples of a printed photo attack[12]

To bypass some of these problems, attackers have thought up a slightly more advanced version of this attack. It consists of the production of a 2D mask. The masks can be done in many ways: they can be simply printed and worn or can consist of different pieces representing different parts of the face of the target. The benefit of this second approach is that the mask will have shapes more like a human face, with different level of depth and shadows. The attacker can even cut holes for their eyes to enable blinking which will fool even some fairly advanced systems which have different levels of liveness detection techniques.

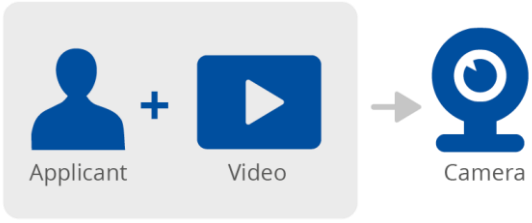
Another way to carry out this type of attack consists in taking possession of a high-quality photo of the victim and showing it through a high-resolution screen, as presented earlier in text.

The cost of the attack is negligible. The attacker only needs a device with a screen or printer to print the victim's photo. Nowadays pretty much anyone has access to a high quality colour printer. Also as regards the screen of a device we all have at least one and in this case the cost can rise significantly only if you want to use a high-resolution screen. The hardest part is to find a good picture to use. However, social media are a good source of data like this and so the cost is reduced to only a matter of time spent on the internet doing some image research.

3.2 VIDEO REPLAY ATTACK

Video Replay attack is an evolution of photo attack replay on screen but with different modalities.

Figure 13: Video replay attack diagram



The attacker will place the screen of his or her device in front of the camera. In this way they will try to fool the system into making it think that the face seen is the one to identify. In this case though, the attacker would have to mind the process used to make the remote identity proofing. If the liveness detection is active and requires some specific facial movements to pass liveness, the video would need to simulate these behaviours.

[12] Bok, J.Y.; Suh, K.H.; Lee, E.C. – *Verifying the Effectiveness of New Face Spoofing DB with Capture Angle and Distance.* Electronics 2020, 9, 66

Figure 14: Samples of a video replay attack[13]

The attack is cheap and easily reproducible. In fact, nowadays everyone has at least a smartphone, meaning that the technology cost is *de facto* erased. The cost and the relative likelihood of success grows by using 4K HDR displays. The difficulty is given by the necessity to obtain a specific video of a specific person to respect the system requirements. However, if the attacker doesn't have access to a truly valid source of tapes of the target and doesn't have the ability to create a valid video to use against the system, this method turns out to be impractical.

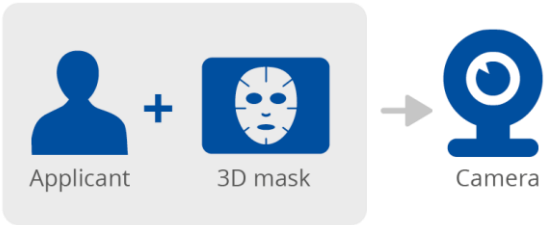
Both photo and video attacks can be delivered through video injection, a distinct attack vector focused on the identity proofing system itself. By leveraging software that allows them to inject the content directly into the capture device, attackers avoid the related security checks.

All kind of systems (fully automated, operator based, hybrid) can require specific tasks to be accomplished, like closing eyes or rotate the head. If the attacker knows the required process, he or she could try to simulate all the required tasks during the creation of the content. However, if the requests are made in real time in a video-based solution, and there is randomness to it, it will be much harder to fool the system.

3.3 3D MASK ATTACK

3D Mask is the most advanced among the attacks that use physical objects. 3D Masks are crafted to reproduce the real traits of a human face. These masks are impressively like human faces and even have eyes holes to fool liveness detection based on eyes gaze, blinking and motion, because the eyes shown through the mask are actual human eyes.

Figure 15: 3D mask attack diagram



3D masks are some of the most difficult to detect for most liveness check systems. Wearing a high fidelity mask, even a human can be fooled in a face-to-face setting. However, they still suffer from the same problems of every other masks. Even if the realism is extreme, the material used to produce the masks has different characteristics with respect to human skin (structure, elasticity/flexibility, blood pulsation, colours, imperfections, etc.). These differences can be leveraged by advanced systems to detect the attack. It is also hard to make a mask that

[13] Bok, J.Y.; Suh, K.H.; Lee, E.C. – *Verifying the Effectiveness of New Face Spoofing DB with Capture Angle and Distance.* Electronics 2020, 9, 66

looks like a specific person – that requires significant skill and is extremely expensive and time consuming even for the best craftsmen.

Figure 16: Example of 3D masks[14]

A variant of this attack is represented using 3D masks that allow the projection of video images on to the mask or that present real displays in the shape of a face. Compared to the other type, it is an attack that is more difficult to perform and has almost all the difficulties of video attacks.

This attack is expensive when the quality of the mask is very high, at par with the special effects used in Hollywood movies. Some pre-made masks based on aspects of random people are to be found on the market. The cost of such a mask is about 3000€, and it does not even represent a specific target. There are companies offering tailored masks based on the features of a specific scanned face. However, it is difficult for an attacker to obtain such a level of detail of a specific target without the consent of his or her victim.

3.4 DEEPFAKE ATTACK

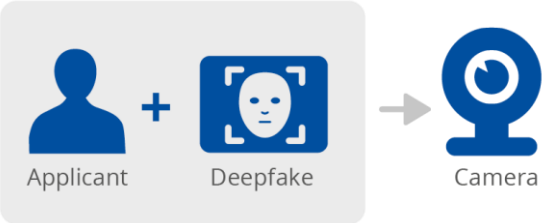
Deepfake software can create a synthetic video or image that realistically represents anyone in the world even if they were never actually performed that action or uttered that phrase. Powerful video processing and augmented reality leveraging machine learning has brought the chance to create deepfake of anyone, without the necessity of having a deep knowledge on video making and special effects.

It could be considered to be in the video replay attack category but given its particularity and ever-increasing importance it is worth talking about it separately.

This technique requires a wide dataset containing images or a video of the target person that the attacker wants to recreate. These data are fed to a program that will learn the fundamental traits of the target and will use the information learned to modify a photo or existing video and apply these traits on the face of the original protagonist – the victim.

[14] Ming, Z.; Visani, M.; Luqman, M.M.; Burie, J.-C. – *A Survey on Anti-Spoofing Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices.* J. Imaging 2020, 6, 139

Figure 17: Deepfake attack diagram



This kind of attack is particularly used against the video evidence-based identity proofing system. If the attacker knows all the steps of the process and can inject the video or present the video on a screen, he or she can fool both an automated system and a system which uses an operator. However, it is worth noting that deepfake attacks have only two ways of getting into the system: either by being presented to the camera or by being directly injected into the camera flow.

A distinct variant of deepfake uses an interactive 3D digital puppet that can be controlled at will starting from one or more photos (face re-enactment). This can be achieved using dedicated software or using software for creating three-dimensional models. With the right software and skills, attackers can achieve results that border on reality and are incredibly lifelike.

Deepfake videos and interactive deepfake puppets can now be created in real-time. In this case the attacker will use his or her camera to capture the video, and in real time the algorithm will modify it with the face of the target it learned and present it to the system. Deepfake can even simulate a different voice. This process was mentioned because, in this way, the operator will effectively talk with a real person, so every randomness added to the process could be ineffective.

Nowadays, such a real time mode is only achievable in high-quality by having a great deal of computing power available and consequently is quite difficult to implement. However, this technological barrier is diminishing – you can already create pretty convincing deepfakes on your smartphone, using one of the popular mobile apps, and this trend will likely continue.

Figure 18: Deepfake example

The difficulty of this attack lies in finding a good dataset of a specific target, especially if he or she is not a famous person. Another factor to keep in mind is the time factor. Most Machine Learning algorithms require hours of training on the specific target to be able to accomplish the task. But we can expect the processing power to continue improving in the future, further lowering the bar for deepfakes.

4. COUNTERMEASURES

The attacks described in the previous chapter can represent a huge problem for remote identity proofing techniques and their application. Therefore companies, researchers and universities have studied them deeply and have devised various systems to face them. Countermeasures can come in different forms, such as extra technical controls, process controls or organisational controls. Every new control can help to mitigate the menace, but one control alone can't prevent them entirely. This is the reason it is a wise decision to apply various controls of different kinds to harden the overall security of the system. In the design and implementation of countermeasures a security-by-design approach should be followed and an analysis of serious risks undertaken.

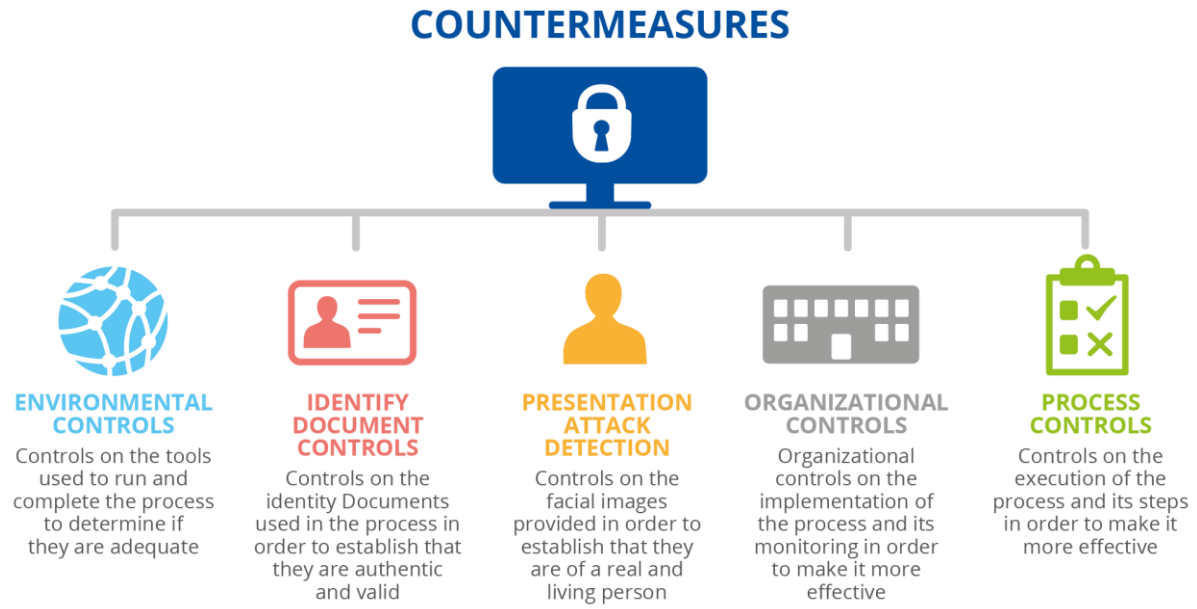
There is no perfect choice of which countermeasures to implement. The best choice is in relation to the type of business, the type and number of users and the degree of assurance that you want to achieve. As with many engineering problems, in choosing countermeasures it is necessary to find the right balance between effectiveness and usability.

It is not possible to reach 100% assurance regarding the effectiveness of countermeasures. The goal is to make the difficulty and cost of executing a fraud higher than the potential benefits.

Countermeasures can be divided into the following main categories:

- environmental controls
- identity document controls
- presentation attack detection
- organisational control
- process controls.

Figure 19: Categories of countermeasures



The opinions expressed in this study are mostly based on our interaction with technology providers and other stakeholders. More biometric lab evaluations and certifications need to be performed in order to provide material proof and verify claims about the effectiveness of various methods.

4.1 ENVIRONMENTAL CONTROLS

Controlling the environment used to carry out the identity proofing process is effective and essential. Environment refers to the hardware, software and network used by the user.

An elementary control to consider is certainly the **verification of the quality level of the video and audio**. This simple check can in fact guarantee that information is received better and consequently guarantee an easier and better proof of identity. In addition, low-quality video can mask attempts to attack, making it difficult to analyse the details of the evidence. To guarantee the quality of the evidence, it is therefore important to define the bandwidth of the connection used and to define a minimum quality level for video and audio to be admitted as evidence.

Another countermeasure is the **execution of the process exclusively through a dedicated application**. On smart devices and computers with webcams, this allows checks to be implemented for the characteristics of a specific device to make sure it is not an emulator, to prevent the injection of video and audio, and to measure the behaviour of the video feed to ensure real-time capture of the feed, interaction with the screen and other variables. On mobile devices, sensors such as an accelerometer might be leveraged. A distinct vein of research focuses on the device and creating an image hash at the very moment of content creation that would carry both time, data and device information and help to spot forgeries[15].

Real fraud story

Abnormal traffic associated with a car dealership located in the United States. In addition to the store employees opening crypto accounts and verifying on the spot in the store, it appeared that the store customers were also asked to pose with their documents as part of the "store procedure". This way free crypto accounts were created on their behalf without their knowledge.

An effective way to identify fraud is to **look for patterns and repetitions** as once an attacker finds a flaw in the process, they tend to exploit it multiple times. For example, an attacker might try to perform the remote identity proofing process several times using the same face but different identities. An effective countermeasure in this sense is **checking the metadata of the remote identity proofing sessions**, such as geolocation, IP, timestamps and time distribution, VPN use and other clues, to look for patterns and indicators of fraud.

4.2 IDENTITY DOCUMENT CONTROLS

To be successful, remote identity proofing needs effective controls to make sure the identity document provided is authentic and has not been lost or stolen, and is not being re-used without the legitimate user's authorisation.

Even if attacks related to the identity document are not within the scope of this report, the related countermeasures that can be implemented cannot be overlooked since the identity

[15] Y. Zheng, Y. Cao, Yuan & Chang, C. Chang – *A PUF-Based Data-Device Hash for Tampered Image Detection and Source Camera Identification,* Nanyang Technological University, 2019

document is a fundamental piece of evidence for the current methods for remote identity proofing.

In the current state of technology, **the highest level of guarantee in terms of an identity document is represented by the electronic identity document equipped with an NFC chip**. The NFC chip contains the document data encrypted and digitally signed by the issuing state. Thanks to the digital signature it is possible to check if the data has been tampered with. Furthermore, by accessing the data stored in the document's NFC chip, it is possible for some ID documents to use the original high-resolution photo in a digital format that guarantees a better quality compared to the paper printed photo on traditional documents and use it to perform more accurate comparisons of faces. However, service providers will no longer have the ability to access this data as access will only be allowed to the police and other authorised bodies.

Where an electronic identity document is not a viable option, optical methods can be used, be they photo-based or video-based. **In terms of security, video based solutions provide more data for analysis and therefore higher assurance that an identity document is authentic.** The use of modern identity documents with multiple security features – such as holograms, watermarks, fine-line guilloche elements, UV prints, micro-texts and others – further raises the assurance level of the remote identity proofing process. However, these security features vary from country to country and from one type of document to another and can be altered more easily than the NFC chip.

Deepfake attacks can be used to forge video-based identity documents in real time. Software can apply the security features typical of a specific document to a simple piece of paper injecting an overlay in real-time video, complete with reflections and the correct management of overlaps.

Also, authentic documents can be used to carry out an attack and, therefore, to implement a process with a high level of assurance. Thus it is essential to be able to **verify that a document is not lost, stolen or expired.** To accomplish this verification step, it is necessary to consult National and International Databases like the *Stolen and lost travel documents database (SLTD- ICAO)*[16]*.* Of course, this check only mitigates the risk and cannot rule it out altogether, as a lost or stolen document might not be reported as such. Also, such databases are not publicly available and some service providers, being private companies, encounter obstacles in accessing them.

To increase the security of identity documents, it is also possible to use digital signatures visible on the document in order to certify its authenticity and obtain useful information for verification such as validity. An example is the new French document which on the reverse side contains a QR code with this function.

Current standards relating to travel documents[17] consider only the biometric data of the face as mandatory. Biometric fingerprint and iris data are optional but would further increase document security. However recent EU legislation on identity cards limits access to such data only to authorised bodies such as police or border guards.

Companies could of course only accept and use documents that comply with the countermeasures listed above but in many real use cases (in relation to the service offered) it is not possible and consequently they are forced to lower the degree of trust.

[16] Interpol - *SLTD database (travel and identity documents)*, https://www.interpol.int/en/How-we-work/Databases/SLTD-database-travel-and-identity-documents
[17] ICAO – *Machine Readable Travel Documents*, 2021, https://www.icao.int/publications/Documents/9303_p9_cons_en.pdf

To improve this part of the process, therefore, **governments must adapt and ensure that the identity documents issued by them meet the highest security standards**. With the proliferation of NFC enabled smartphones, the rollout of electronic identity documents is likely to continue.

Going beyond verifying the authenticity of the document and considering the next part of the remote identity proofing process, namely face comparison, there are other countermeasures that governments could take to improve security and effectiveness.

Considering the good practices analysed in the domain area, it could be useful to **establish stringent rules regarding the type of photo required to issue the identification document.** Some European countries, in fact, have established that the photos must only be in digital format to check for any manipulation as manipulation can be more easily verified in this format.

Another countermeasure could be to acquire the photos only on site through dedicated staff or the use of certified photographers who provide the digitally signed photo to the applicant or send it directly to the agency. It could also be useful to acquire facial photos from different angles and save them on the NFC chip (compatible with the memory capacity) to have more data to perform the comparison.

In addition to using the controls illustrated, the following controls can be implemented by companies to improve the verification of the authenticity of the identity document:

- define the list of identity documents that are allowed for the process
- define the acceptance criteria for the validity of the different identity documents.

4.3 DETECTION OF PRESENTATION ATTACKS

The core of the automated processes is represented by the software used to perform the **Presentation Attack Detection (PAD)**. These software systems use artificial intelligence and machine learning to understand whether a face is that of a living and real person and to do so they try to verify certain characteristics. Their operation therefore varies according to the characteristics verified.

Liveness techniques focus on some aspects that are intrinsic in human being. There are different levels of hardening, which imply that different technologies be included in the system.

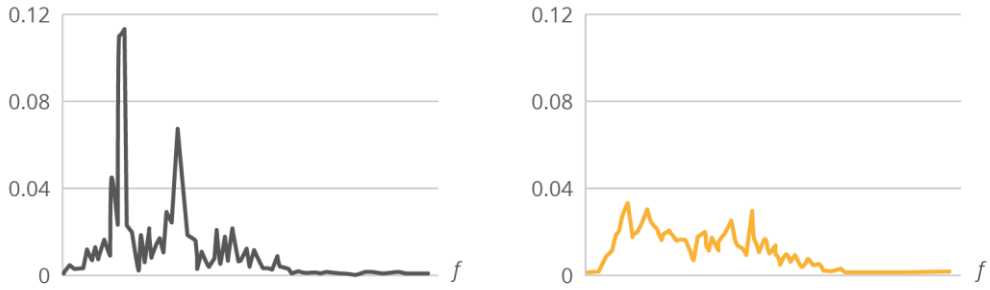
Active controls require user action. The request for random movements, in addition to verifying that the video is not recorded, is considered effective against unsophisticated deepfake attacks, particularly if the user is **asked to perform fast movements** as the software or the underlying processing power struggles to keep up. It is also considered effective to **ask the user to place a hand or an object in front of the face** as the software struggles to manage the overlap. Another way to achieve this is by **checking the motion of the eyes following an object on screen**. This is not something that an attacker can forecast with a recorded video, so again randomness helps to protect from malicious attacks.

Figure 20: Example of imperfections in deepfake caused by rapid movements[18]

Passive controls require no user action and focus on involuntary human signals present in the analysed images or video frames such as **facial expression, face warping artefacts** and **resolution inconsistencies** resulting from a deepfake production pipeline[19], **unnatural or missing blinking, poorly generated eyes or teeth, pupil dilatation and other variables[20]. Random colours can be flashed on the screen** to analyse reflections on the subject and their surroundings, even though some of the stakeholders consulted claim it is possible to generate semi-transparent colours instantly on the deepfake puppet, defeating this liveness method.

A more advanced technique can instead **check for microvariations in the intensity of skin colour given by blood pulse**. In fact, even if to recognise these changes is tricky for human eyes, a machine learning model can focus on these aspects and recognise them. A physical reproduction of a face is usually not able to represent blood pulse, and re-creating the effect in a video is also not trivial, meaning that it represents a discriminating factor. The detection can be achieved with ML models that provide different layers of analysis of the face presented, by exploiting the study of images obtained by RGB cameras.

Figure 21: Typical power spectrum distribution patterns of a real access (left) and a mask attack (right) extracted from the green colour channel[21]

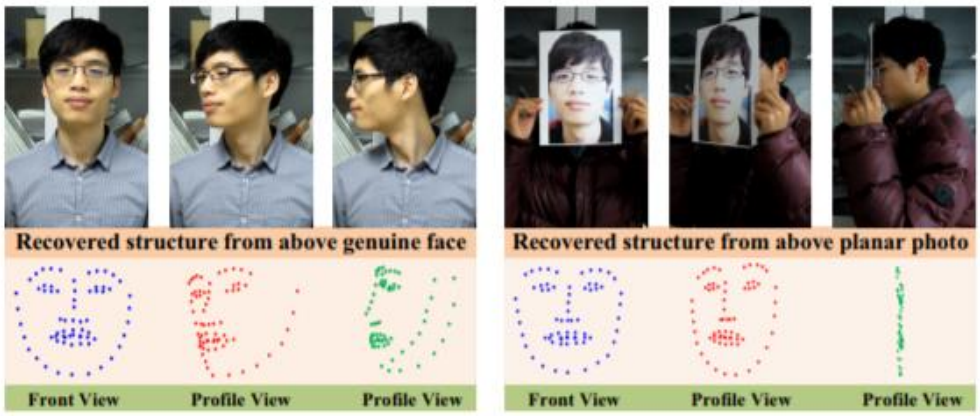


The **surface** of a person's skin has certain characteristics when it reacts to light and these can vary from person to person. A mask made of artificial material hardly reacts to light and does not reflect it in the same way as human skin. Even though make-up can suppress the difference, using an intensity distribution function to describe the difference in the reflectance of the images makes it possible to detect whether the person in front of the system is wearing a mask or not.

[18] Ubble.ai – *Création d'une attaque par altération numérique du visage,* France, 2021
[19] Y. Li, S. Lyu – *Exposing DeepFake Videos By Detecting Face Warping Artifacts,* University of Albany, The State University of New York, 2018
[20] F. Matern, C. Riess, M. Stamminger – *Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations,* Nuremberg, 2019
[21] X. Li, J. Komulainen, G. Zhao, P. Yuen, M. Pietik – *Generalized face anti-spoofing by detecting pulse from face videos,* University of Oulu, Oulu, Finland, December 2016

Three-dimensional methods to analyse face depth produce 3D maps of a user's features, checking on the depth and geometry of their face. With 3D liveness technology the number of variables and involuntary human liveness traits that can be considered in the proofing process is significantly increased compared to 2D or previous methods. This makes 3D liveness solutions more reliable in terms of accuracy, usability and the prevention of attacks.

Figure 22: A comparison of recovered sparse 3D facial structures between genuine and photo face, showing significant differences[22]



Verifying that the user's face exists in three dimensions can be done through dedicated time-of-flight cameras such as laser-based lidars (that resolve distance from the object by measuring round trip time of a light signal), infrared light and other sensors, or through ubiquitous 2D cameras that acquire images at different distances and from different angles. Dedicated 3D sensors can collect more information faster and with less interaction from the user, but they are not yet widespread and so 3D face maps generated from reverse engineering images from a 2D camera are the most viable short-term option nowadays.

Figure 23: 3D camera vs classic cameras diagram

Using multiple cue-based techniques allows the system to gather more information. Since remote identity proofing is a decision-making process, it becomes easier by having more information to use. For this reason, a **simultaneous use of multiple technologies in the same system would be recommended where possible,** provided it does not reduce usability or accessibility. For example, the process could check the texture of cue-based evidence to see if a mask was being presented and reinforce the process by checking for facial movement. This can increase the chance of detecting an attack. New technologies are also emerging in the field of software that perform presentation attack detection, but any that require specialised hardware are not likely to be used in the near term.

Even if PAD is widely used as the main countermeasure, it is important to underline the importance that the human factor still has. **There are cases, such as coercion or new types of attack, where the human operator is more effective than software.** In this regard, many

[22] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li – Face *liveness detection using 3d structure recovered from a single camera,* in *IAPR International Conference on Biometrics*, ICB, June 2013

providers use mixed approaches where the human operator intervenes when the software is unable to decide. To do this, assurance thresholds can be set on the software, beyond which operator intervention is required.

PAD software is based on neural networks and machine learning algorithms and for this reason it is essential that the **data in the database used to train these algorithms are consistent with the purpose of the algorithm and are labelled correctly**. Furthermore, such databases must be constantly **updated to mitigate the trends** in the **latest attacks** and the conclusions taken by the algorithm must be verified to ensure their good functioning.

4.4 ORGANISATIONAL CONTROLS

Besides considering the implementation details of the process for remote identity proofing, general and organisational countermeasures also need to be implemented.

The first is certainly to **follow industry standards if available** and, in the fast changing landscape of biometric cybersecurity, stay up-to-date with the latest threat vectors and factor in the limits of older standards.

In relation to remote identity proofing, ISO/IEC 30107 establishes principles for evaluating the performance of methods used for the detection of presentation attacks. Of course, the attack vectors are constantly changing and consequently ISO/IEC 27001 provides requirements for an organisation's management system for information security. Other standards from the ISO/IEC 27000 family are also relevant.

Since attack scenarios are rapidly evolving, all the more so in the digital world, we can expect more standards, certifications and testing methods to emerge. A potential new certification scheme for remote identity proofing should encompass different levels of assurance and tackle both face protection attacks and the verification of identity documents.

Remote identity proofing methods were also analysed by a special task force of the European Telecommunications Standards Institute (ETSI)[23]. The standards developed are:

- ETSI TR 119 460 *Electronic Signature and Infrastructures (ESI); Survey of technologies and regulatory requirements for identity proofing for trust service subjects*
- ETSI TS 119 461 *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for identity proofing of trust service subjects, including for remote processes.*

The latter fills the gap in former European standards published by ETSI on trust services that only specify identity proofing by generic requirements such as 'physical presence or means which provide equivalent assurance as physical presence' derived from eIDAS art. 24.1, by providing concrete and measurable requirements including for remote processes.

If involved in the process, human operators are a fundamental element for the effectiveness of the result, and it is necessary to ensure that their task can be performed easily and is monitored. In this regard, the following controls can be implemented:

- allowing the operator to stop and void the remote identity proofing should any suspicion come to his or her mind, without the need to provide any justification to the applicant;

[23] ETSI – *Specialist Task Force 588: Identity Proofing for Trust Service Subjects*, https://portal.etsi.org/STF/STFs/STF-HomePages/STF588

- assigning a particular registration officer for a specific remote identity proofing process should not be predictable;
- providing proper and continuous training for the operators, focused on both their role in identity proofing and on social engineering attacks that might push them to circumvent the controls;
- defining and implementing a monitoring process;
- providing a secure and well-organised workplace for the operator.

Another countermeasure that can be implemented is that of **designing a linear and understandable process** to guarantee good performance by the users and operators. A complex process could result in a loss of awareness about the actions being performed, the expected result for those actions, and why they are being performed.

A very effective control that can give real and practical results consists in the **creation of a bounty program** that provides financial rewards to those who manage to evade the controls of a remote identity proofing process and identify themselves with another identity or fool the system into saying that a spoof artifact or video is a real, live human.

It is also important to use a **risk-based approach** and use a robust **risk analysis** methodology aligned with best practices, to identify current threats but, above all, future and unknown ones.

There is a need for governments and other institutions to play their part as well, for example by launching **external, objective and impartial test frameworks for the services available on the market.** A good example of this approach is the US National Institute of Standards and Technology (NIST) running their 2D *Face Recognition Vendor Test*[24]. However, according to many of the stakeholders we consulted, what the industry is lacking today is an established way to test the performance of methods to detect liveness. Also, the industry lacks a clear framework for the type of testing needed to determine whether different approaches to injection attacks can be effectively deterred.

4.5 PROCESS CONTROLS

Controls can be implemented on the execution of the process and its steps to make it more effective. It is important to implement a periodic testing and monitoring of the performance of the entire system for remote identity proofing (software, human operators, etc.). In addition, self-inflicted attacks should be attempted to evaluate the system's reliability in detection (breakthrough tests).

Following a security by design approach allows to reduce exposure to possible attacks to be reduced. In this regard, the following countermeasures can be implemented:

- define a supplemental list of evidence to strengthen the process, manage extreme cases or when doubts arise (e.g. ask for a photo of an energy bill or bank account balance as additional evidence);
- define the rights and obligations of all participants and parties who rely on the process;
- request that the remote identity proofing process happens in real time when it requires the participation of the subject but allow for asynchronous evaluation of its trustworthiness;
- check the behavioural patterns of the subject to verify that he or she is not coerced, is acting voluntarily and understands the implications of the process;
- record the session content (video, pictures, audio) and metadata and store it in a tamper proof way (in compliance with the General Data Protection Regulation);

[24] NIST – *Face Recognition Vendor Test (FRVT),*
https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-ongoing

- require active participation of the applicant, including some speech;
- introduce random elements like changing the order of the questions or asking the applicant unexpectedly to raise a hand, rotate the face and similar gestures, varying these processes from time to time, as scammers adjust quickly to established routines;
- use breakthrough testing that considers all ways to penetrate the system, whether it is to exploit software flaws or to deceive the software or the human making the decision;
- duplicate the process by handling cases in parallel to verify that they return the same result.

5. CONCLUSIONS

The boom in digital services is changing consumer preferences, raising the demand for greater convenience and better user experiences. Today, customers expect access to services anytime and anywhere, hence the exponential success of digital and mobile services. **The need to securely onboard and prove a customer's identity remotely is therefore becoming critical for organisations providing such a service.** Organisations that have already implemented digital onboarding and remote identity proofing solutions have seen the benefits and challenges.

Criminals are creative in devising tactics to circumvent vulnerabilities in a company's cybersecurity. Even during the pandemic, they have continued to find ways to infiltrate systems and exploit services. Furthermore, since there are several types of identity documents, there are many possible threats. **Given the circumstances surrounding COVID-19, fraudsters have also been quick to capitalise on this drastic change in circumstances, finding new opportunities to commit fraud.** In recent years, the number of attacks with masks and video injection has been increasing.

As we have seen, the remote identity proofing process is already of fundamental importance, and in the future it will be increasingly so because the world is moving in that direction. Underlining this importance is the work of ETSI[25], a regional standards body dealing with telecommunications, broadcasting and other electronic communications networks and services. ETSI created a specialist task force[26] that has produced specifications on identity proofing for trust services as defined by eIDAS[27]. eIDAS is also focusing more on remote identity proofing and a planned revision aims for a further development of trust services and eID in Europe.

To fight back against possible threats, we must first be aware of them as well as understand the tools and methods available to attackers. It is not only important to protect the verification of biometric data; the whole process needs to be secured from point to point. **We need to analyse threats, employ a risk-based approach to prioritise our countermeasures, and always be on top of new trends to stay ahead of attackers.** Spoof and camera bypass bounty programmes can also help to uncover emerging threat vectors and help develop mitigation measures that raise the cost of attacks and reduce their feasibility.

In the wild, identity and technology providers have implemented both active and passive security controls which mostly involve the use of video and operator intervention. The video helps to perform a greater number of security checks and the operator helps the artificial intelligence to identify any new types of attacks.

Today's remote identification systems are mainly based on facial recognition. Therefore, the implementation of artificial intelligence and detection algorithms is focusing on biometric data. **Although many have faith in technology and believe that having a fully automated process without human intervention will not be long in coming, humans are yet to remain in the loop.** Quoting the German philosopher Feuerbach, 'we are what we eat', the same concept applies to algorithms and artificial intelligence. Algorithms cannot understand and detect new fraud on their own and human action is required to assign the correct labels to new

[25] European Telecommunications Standards Institute (ETSI), https://www.etsi.org/
[26] ETSI – *Specialist Task Force 588: Identity Proofing for Trust Service Subjects*, https://portal.etsi.org/STF/STFs/STF-HomePages/STF588
[27] European Commission – *eIDAS Regulation, https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation*

attacks. Therefore, humans are needed to clean and tag data enabling high quality training that will result in better performances and the mitigation of adversarial attacks.

According to this study's findings, the main attacks on remote identity proofing systems today are low-tech, low-probability of success attacks. But the last frontier can be considered the deepfake attack presented via video injection or high fidelity screens. **Deepfakes are a real threat to the future and will not only have an impact on the remote identification industry but will also affect the press, telecommunications and other fields.**

To secure the process of remote identity proofing, there is a need to follow a security-by-design approach and a holistic security perspective that comprehends all phases of the identification cycle. A serious risk analysis will be necessary because it all depends on the level of risk which is closely related to the abilities and incentives of the attackers. Then, policymakers may weigh in to set the minimal security thresholds for given user scenarios. The key is to make the attackers' job more difficult, time-consuming and expensive.

The future of attacks is a complex issue. We hope this report will benefit continuous structured efforts in risk analysis in this field and contribute to the modulation of countermeasures, helping remote identity proofing to remain trustworthy and reliable in the years to come.

6. BIBLIOGRAPHY & REFERENCES

6.1 BIBLIOGRAPHY

- X. Li, J. Komulainen, G. Zhao, P. Yuen, M. Pietik, *Generalized face anti-spoofing by detecting pulse from face videos*, University of Oulu, Oulu, Finland, December 2016.
- J.Y. Bok, K.H. Suh, E.C. Lee, *Verifying the Effectiveness of New Face Spoofing DB with Capture Angle and Distance*. Electronics 2020, 9, 66.
- Z. Ming, M. Visani, M.M. Luqman, J.-C. Burie, *A Survey on Anti-Spoofing Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices*. J. Imaging 2020, 6, 139.
- T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, *Face liveness detection using 3d structure recovered from a single camera*, in IAPR International Conference on Biometrics, ICB, June 2013.
- A. Bashir, Y.A. Fadlalla, *Techniques of Detecting Forgery in Identity Documents*, IEEE, Elnihood, Sudan November 2017.
- A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos*, What does presentation attack detection and liveness actually mean?* Biometrics Institute Limited, Imperial House, 8 Kean Street London WC2B 4AS, UK.
- C. Busch, *What is a Presentation Attack? And how do we detect it?,* Dan Panorama, Tel Aviv, January 16, 2018*.*
- Christoph Busch, *Presentation Attack Detection - ISO/IEC 30107,* NTNU – Norwegian University of Science and TechnologySeptember 2020.
- IDnow, *Trend Report: Identity Fraud 2021. How to win the cyber fraud arms race*, Munich, Germany, 2021.
- J. Hernandez-Ortega, J. Fierrez, J. Galbally, A. Morales*, Introduction to Face Presentation Attack Detection,* Springer, April 2019.
- L. Lia, Z. Xiaa, X. Jianga, Y. Maa, F. Roli, X. Fenga, *3D Face Mask Presentation Attack Detection Based on Intrinsic Image Analysis.,* School of Electronics and Information, Northwestern Polytechnical University, Xi'an, Shaanxi, China and Department of Electrical and Electronic Engineering, University of Cagliari, Cagliari, Sardinia, Italy, March 2019.
- L. Zhao, C. Chen, J. Huang, *Deep Learning-based Forgery Attack on Document Images,* The Institute of Electrical and Electronics Engineers (IEEE), February 2021.
- M. Fang, N. Damer, F. Kirchbuchner, A. Kuijper, *Real Masks and Spoof Faces: On the Masked Face Presentation Attack Detection*, Fraunhofer Institute for Computer Graphics Research IGD and Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany, March 2021.
- M. Gomez-Barrero, *Presentation Attack Detection and Unknown Attacks,* IFPC*,* 2020 October.

- Meng Shen, Yaqian Wei, Zelin Liao, and Liehuang Zhu. 2021. *IriTrack: Face Presentation Attack Detection Using Iris Tracking.* Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 5, 2, Article 78 (June 2021), 21 pages.
- Z. Ming, M. Visani, M.M. Luqman, J.-C. Burie, *A Survey on Anti-Spoofing Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices*. J. Imaging 2020, 6, 139.
- N. Erdogmus and S. Marcel, *Spoofing Face Recognition With 3D Masks, in IEEE* Transactions on Information Forensics and Security, vol. 9, no. 7, pp. 1084-1097, July 2014.
- N. Kose and J. Dugelay, *On the vulnerability of face recognition systems to spoofing mask attacks*, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 2357-2361.
- P. Korshunov, Idiap Research Institute, *Deepfake detection: humans vs machines*, Martigny, Switzerland.
- P. Grother, M. Ngan, K. Hanaoka, *Ongoing Face Recognition Vendor Test (FRVT*), National Institute of standards and technology (NIST), United States of America, 2021/09/10.
- R. Ramachandra, C. Busch, *Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey*, A CM Computing Surveys, Volume 50 Issue, 1 April 2017, Article No.: 8pp 1 - 37.
- R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, *DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection*, Biometrics and Data Pattern Analytics - BiDA Lab, Universidad Autonoma de Madrid, Spain, June 2020.
- S. Bhattacharjee, A. Mohammadi and S. Marcel, *Spoofing Deep Face Recognition with Custom Silicone Masks*, 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), 2018, pp. 1 - 7.
- S. Jia, G. Guo, Z. Xu, Q. Wang*, Face presentation attack detection in mobile scenarios: A comprehensive evaluation*, Image and Vision Computing, Volume 93, 2020, 103826, ISSN 0262-8856.
- S. Kumar, S. Singh, J. Kumar, *A Comparative Study on Face Spoofing Attacks,* IEEE 2017 May, Greater Noida, India.
- S. Marcel, *Biometric Face Presentation Attack Detection with Multi-Channel Convolutional Neural Network*, IEEE Transactions on Information Forensics and Security, 2019.
- Sanders, J.G., Ueda, Y., Yoshikawa, S. et al. *More human than human: a Turing test for photographed faces*. Cognitive Research 4, 43 (2019).
- U. A. Ciftci, I. Demir, L. Yin, S. Member, *FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals,* IEEE Transactions on Pattern Analysis and Machine Intelligence*,* July 2020.
- Ubble.ai, *Remote identity proofing, other identification methods, risk analysis and recommended mitigations to reach a high level of confidence*, Paris, France, 2021.
- Ubble.ai, *Création d'une attaque par altération numérique du visage*, Paris, France, 2021.
- Veriff, *Fraud report 2020*, Tallinn, Estonia, 2020.
- Y. S. El-Din, M. N. Moustafa, H. Mahdi, *Adversarial Unsupervised Domain Adaptation Guided with Deep Clustering for Face Presentation Attack Detection*, Computer and

Systems Engineering Department, Ain Shams University, Cairo and Department of Computer Science and Engineering, The American University in Cairo, New Cairo, Egypt, February 2021.

- Yang, L., Song, Q. & Wu, Y. *Attacks on state-of-the-art face recognition using attentional adversarial attack generative network*. Multimedia Tools and Applciations 80, 855–875, 2021.
- Y. Li, S. Lyu – *Exposing DeepFake Videos By Detecting Face Warping Artifacts*, University of Albany, The State University of New York, 2018.

6.2 ENISA Publications

ID	Description
Remote ID proofing	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
eIDAS Compliant eID Solutions	*eIDAS Compliant eID Solutions, Security Considerations and the Role of ENISA*, March 2020 https://www.enisa.europa.eu/publications/eidas-compliant-eid-solutions
ENISA Security Framework for TSPs	*Security Framework for Trust Providers*, March 2021 https://www.enisa.europa.eu/publications/security-framework-for-trust-providers/

6.3 Applicable Legislation / Regulation

ID	Description
AMLD5	Directive (EU) 2018/843 of the European Parliament and of the Council amending directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing ('AMLD5') https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L0843
CIR 2015/1501	Commission Implementing Regulation (EU) 2015/1501 of 8 September 2015 on the interoperability framework pursuant to Article 12(8) of Regulation (EU) No 910/2014 of the European Parliament and of the Council on electronic identification and trust services for electronic transactions in the internal market. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:JOL_2015_235_R_0001
CIR 2015/1502	Commission Implementing Regulation (EU) 2015/1502 of 8 September 2015 on setting out minimum technical specifications and procedures for assurance levels for electronic identification means pursuant to Article 8(3) of Regulation (EU) No 910/2014 of the European Parliament and of the Council on electronic identification and trust services for electronic transactions in the internal market. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R1502

ID	Description
eIDAS	Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive1999/93/EC. OJ L 257, 28.8.2014, p. 73–114. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv%3AOJ.L_.2014.257.01.0073.01.ENG
EU-ID	EU digital ID scheme for online transactions across Europe https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12528-European-Digital-IdentityEUid
FATF-ID-G	The FATF Digital Identity Guidance, issued in March 2020 http://www.fatfgafi.org/publications/financialinclusionandnpoissues/documents/digital-identity-guidance.html
FATF-R	The FATF Recommendations as amended June 2019 http://www.fatf-gafi.org/publications/fatfrecommendations/documents/fatf-recommendations.html
LOA GUIDE	Guidance of the European Cooperation Network on the application of the levels of assurance which support the eIDAS Regulation https://ec.europa.eu/cefdigital/wiki/download/attachments/40044784/Guidance%20on%20Levels%20of%20Assurance.docx
SIDCR	Regulation (EU) 2019/1157 of the European Parliament and of the Council of 20 June 2019 on of Union citizens and of residence documents issued to Union citizens and their family members exercising their right of free movement https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32019R1157
A/CN.9/WG.IV/WP.164 – UNCITRAL CROSS DRAFT	UNCITRAL Draft Provisions on the Use and Cross-border Recognition of Identity Management and Trust Services – synthesis of comments submitted by States and international organisations https://undocs.org/en/A/CN.9/WG.IV/WP.164
UNCITRAL ECOMMERCE	UNCITRAL Working Group IV: Electronic Commerce https://uncitral.un.org/en/working_groups/4/electronic_commerce
A/CN.9/WG.IV/WP. 162 - UNCITRAL IDM DRAFT	UNCITRAL Draft Provisions on the Use and Cross-border Recognition of Identity Management and Trust Services (A/CN.9/WG.IV/WP.162, April 6-9, 2020) https://undocs.org/en/A/CN.9/WG.IV/WP.162
A/CN.9/WG.IV/WP. 150 - UNCITRAL IDM TC	UNCITRAL Terms and concepts relevant to identity management and trust services, (A/CN.9/WG.IV/WP.150, February 6, 2018) https://undocs.org/en/A/CN.9/WG.IV/WP.150

6.4 Standards and Others

ID	Description
ANSSI PVID	*Prestataires de vérification d'identité à distance - Référentiel d'exigences* - Version 1.0 du 19 Novembre 2020. https://www.ssi.gouv.fr/uploads/2020/11/anssi_pvid_referentiel_exigences-v1.0.pdf
BSI TR-03147	BSI TR-03147 (v1.0.4) *Assurance Level Assessment of Procedures for Identity Verification of Natural Persons.*
CC	Common Methodology for Information Technology Security Evaluation, Common Criteria Portal. https://www.commoncriteriaportal.org/files/ccfiles/CEMV3.1R5.pdf
ETSI TS 119 431-1	ETSI TS 119 431-1 (v1.1.1): *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for trust service providers; Part 1: TSP service components operating a remote QSCD / SCDev.*
ETSI TS 119 431-2	ETSI TS 119 431-2 (v1.1.1): *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for trust service providers; Part 2: TSP service components supporting AdES digital signature creation.*
ETSI TR 119 460	ETSI TR 119 460 (v1.1.1): *Electronic Signatures and Infrastructures (ESI); Survey of technologies and regulatory requirements for identity proofing for trust service subjects.*
ETSI TS 119 461	ETSI TS 119 461: *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for trust service components providing identity proofing of trust service subjects* which will be published after the present report. (for public review before publication)
ETSI EN 319 401	ETSI EN 319 401 (v2.2.1): *Electronic Signatures and Infrastructures (ESI); General Policy Requirements for Trust Service Providers.*
ETSI EN 319 411-1	ETSI EN 319 411-1 (v1.2.2): *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for Trust Service Providers issuing certificates; Part 1: General requirements.*
ETSI EN 319 521	ETSI EN 319 521 (v1.1.1): *Electronic Signatures and Infrastructures (ESI); Policy and security requirements for Electronic Registered Delivery Service Providers.*.
ICAO-SLTD	ICAO, SLTD Project. providing law enforcement with instant worldwide access to INTERPOL databases, https://www.icao.int/Meetings/TAG-MRTD/Documents/Tag-Mrtd-18/Interpol.pdf.
ISO/IEC 15408	ISO/IEC 15408 *Information technology - Security techniques - Evaluation criteria for IT security.*
ISO/IEC 27001	ISO/IEC 27001:2013: *Information technology -- Security techniques -- Information security management systems -- Requirements*.
ISO/IEC TS 29003	ISO/IEC TS 29003:2018: *Information technology — Security techniques — Identity proofing.*

ID	Description
ISO/IEC 30107	ISO/IEC 30107-1:2016 *Information technology — Biometric presentation attack detection — Part 1: Framework.* ISO/IEC 30107-2:2017 *Information technology — Biometric presentation attack detection — Part 2: Data formats.* ISO/IEC 30107-3:2017 *Information technology — Biometric presentation attack detection — Part 3: Testing and reporting.* ISO/IEC 30107-4:2020 *Information technology — Biometric presentation attack detection — Part 4: Profile for testing of mobile devices.*
NIST SP 800-63-3	NIST SP 800-63-3: *Digital Identity Guidelines.*
NIST SP 800-63A	NIST SP 800-63A: *Digital Identity Guidelines; Enrollment and Identity Proofing Requirements.*
PRADO	*PRADO – Public Register of Authentic travel and identity Documents Online.* https://www.consilium.europa.eu/prado/en/prado-start-page.html

A ANNEX: METHODOLOGY

A.1 DESK RESEARCH

Desk research activities were conducted throughout the preparation of this report. The study team collected and examined all publicly available data on the topic, including reports, surveys, publications, conference papers and other sources.

A.2 INTERVIEWS

Another fundamental step of the triangulation methodology was the interview process. A programme of interviews and an online survey targeted stakeholders across all relevant sectors within the scope of the study. Desk research activities were of primary importance in identifying relevant stakeholders for participation in the interview programme.

Stakeholders were selected according to 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. Desk research activities were pivotal in identifying the relevant stakeholders to be engaged.

A.3 SURVEY

The interviews and the online questionnaire are complementary tools which were used to perform and complete the data collection and inform the analysis. Dissemination of the questionnaire was crucial in enabling the validation and further articulation of information obtained through the review of literature as well as identifying and filling in gaps in information. Perhaps the survey was used as an alternative to the interview when stakeholders were not available for it.

The survey comprised a web-based questionnaire published on the EU Survey platform. The structure of the questionnaire had foreseen a total of 18 questions investigating three main topics of interest:

- state of the art and the future of RIDP
- attacks
- countermeasures.

Furthermore, the questions were tailored for three main categories of stakeholders which together comprise all the categories identified in the literature and by desk research:

- technology providers (CAT I)
- organisations using RIDP (CAT II)
- researchers (CAT III)

Questions were either multiple choice, where stakeholders were asked to provide a ranking of different options, or closed-ended questions with just yes/no replies. The survey was designed in tandem with the interview questions allowing most of the time for comments. In fact, the team aimed at ensuring to a certain extent a correspondence between questions so as to ensure comparability between the dataset from the survey and information acknowledged during the interviews.

A.4 WORKSHOP

On July 15, 2021 ENISA and PwC held a workshop *Remote Identity Proofing Practices: Attack Scenarios* where the preliminary findings and draft recommendations of this study were presented to stakeholders, including representatives and experts from academia, technology or service providers and national standardisation authorities. The workshop served as an opportunity for participants to share their experiences, discuss actual cases, and gather feedback.

To meet the objectives set, a restricted panel of relevant speakers was involved. Panellists were selected and engaged according to their expertise in the field of remote identity proofing and their willingness to share experiences and views on these specific topics.

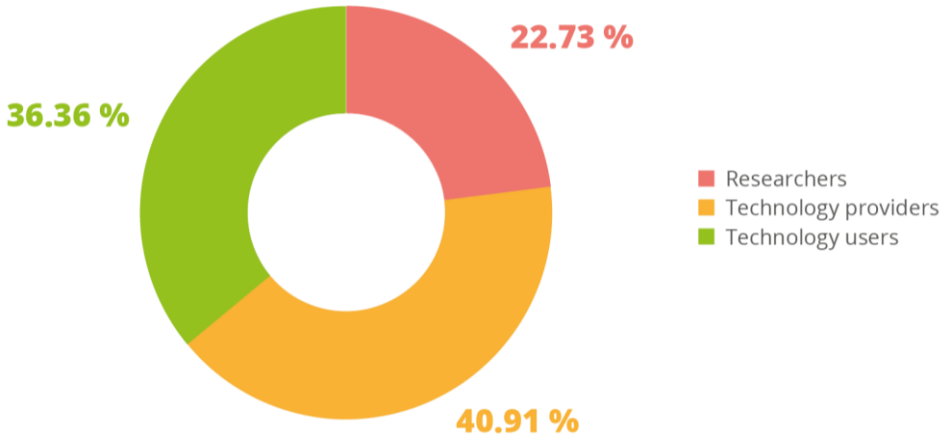
The attack scenarios explored consisted of:

- deepfake video injection
- high-quality 3D silicone masks
- video manipulation of an identity document.

B ANNEX: SURVEY RESULTS

We received feedbacks from a total of 22 stakeholders, of whom 41% belonged to CAT I, 36% represented CAT II and 23% of responses came from CAT III. Given the small sample, the results listed here may not be truly representative of the industry and academia.

Figure 24: Survey share of respondents



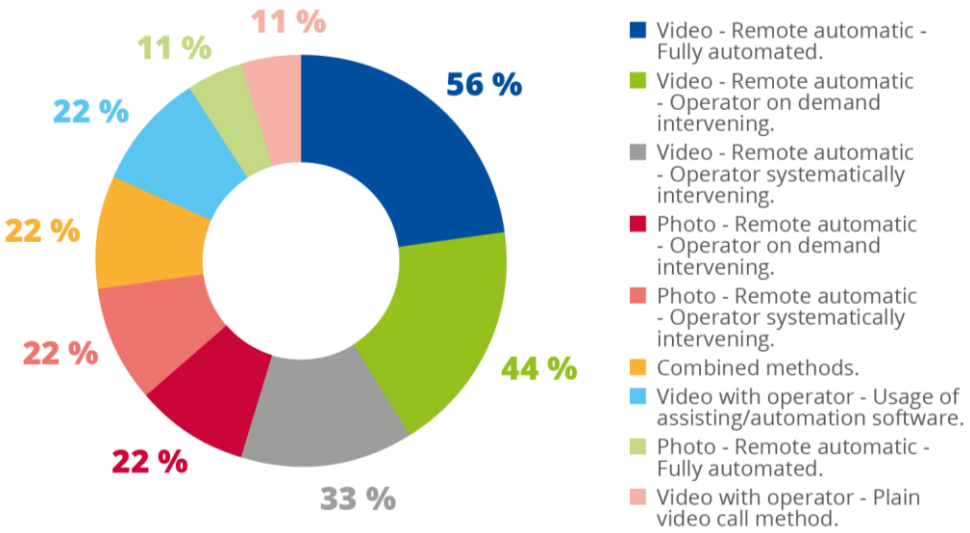
Below, the team provides an analysis of the results of the survey questionnaire divided by the previously described categories.

B.1 SURVEY RESULTS FROM TECHNOLOGY PROVIDERS

With a share of 41%, technology providers is the most represented category within the three identified. In line with the results obtained in the interviews, the survey displays the state of the art of technologies implied in the remote identity proofing process.

Replies related to technology currently in use showed that 56% (5 votes) of stakeholders belonging to this category currently offer 'Video - Remote automatic - Fully automated', with 44.5% (4 votes) 'Electronic identification means' and with a share of 33% (3 votes) 'Video with operator'. Furthermore, regarding technologies, they are gearing up to offer in the near future 'Video – Remote automatic – Fully automated' with 55%, 'Video – Remote automatic – Operator on demand intervening' with 44% and 'Video - Remote automatic - Operator systematically intervening' with 33%.

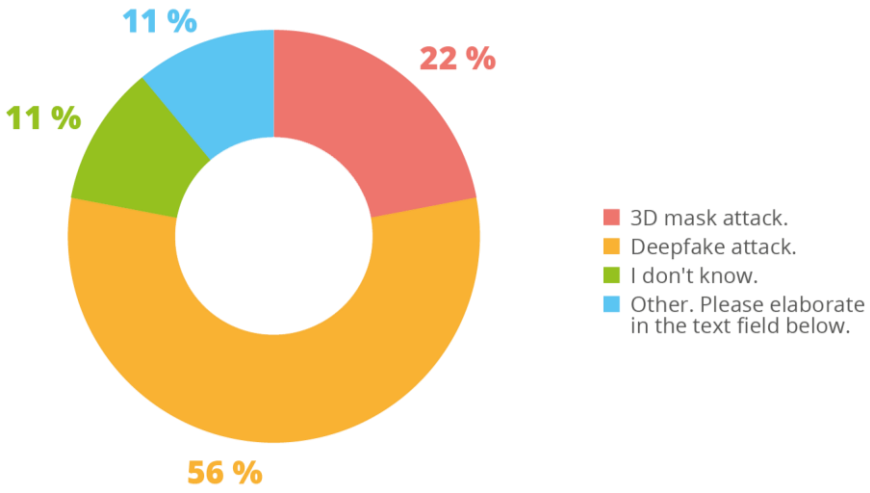
Figure 25: RIDP methods foreseen in the future



Concerning the effectiveness of the possible RIDP methods, an equal percentage of 22% of survey respondents voted for 'Video with operator - Usage of assisting/automation software', 'Electronic identification means' and 'Combined methods.

We then asked which face presentation attack was most likely to succeed; deepfake attacks (33%) and 3D mask attacks (22%) are considered the attacks with the highest probability of success. In addition, the types of attacks just mentioned are also the ones considered most concerning for the future.

Figure 26: Future of RIDP attacks

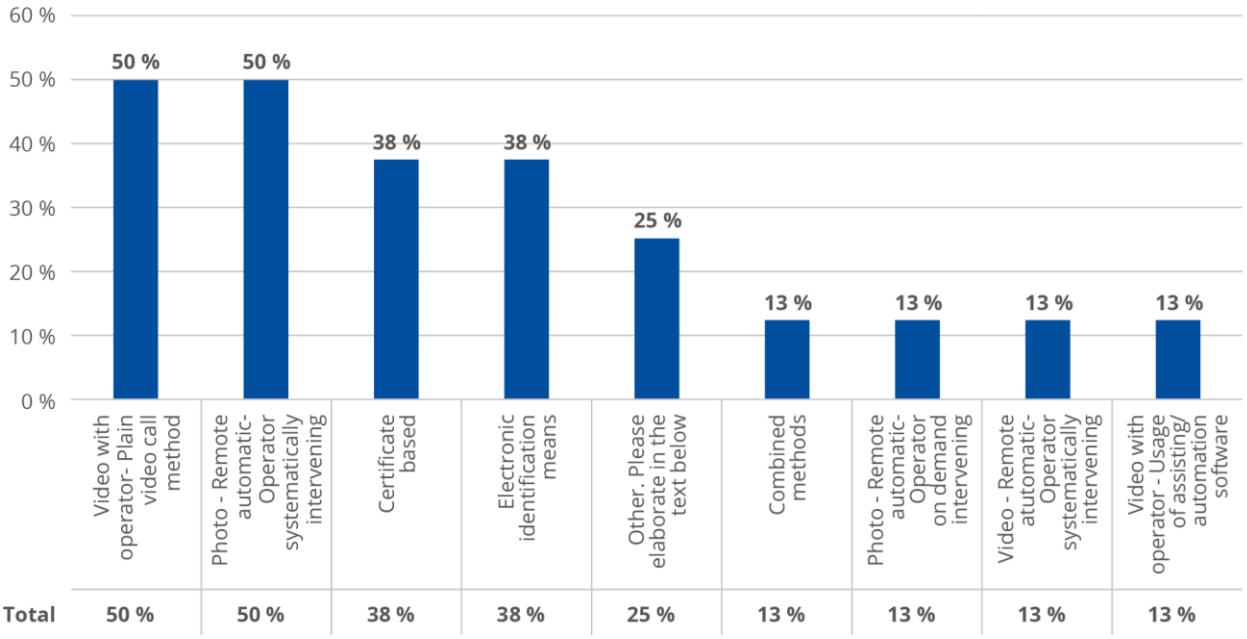


With regards to the deployment of technical controls, the most recommended were 'Detection software - New Technology (cutting-edge machine learning)' and 'Define the acceptance criteria for the validity of the different identity documents' both of which received 55% of the votes. At the same time 'Detection software - Multiple cue-based' together with 'Define a minimum acceptable level of audio and video quality, and Internet bandwidth and latency' scored 44% of the votes from survey participants.

B.2 SURVEY RESULTS FROM ORGANISATIONS USING RIDP TECHNOLOGIES

The graph below displays the answers received from RIDP technology users about the methods they currently use. The data gathered is in line with information obtained in the interviews. In fact, not surprisingly, the methods at 50% are 'Video with operator - Plain video call method' and 'Photo - Remote automatic - Operator systematically intervening'. The two methods just mentioned could involve liveness sessions but in general they are based more on biometric comparisons. By contrast, as survey results show, methods in use in the wild not based on facial recognition such as 'Certificate based, and 'Electronic identification means' have a consistent share of 38%.

Figure 27: RIDP methods currently in use



Regarding the future, the methods foreseen are generally characterised by an approach described as 'technological automation'. Differently, others still believe that in the future there will also be a need to include a human operator in the remote identity proofing processes.

Figure 28: Methods foreseen in the future

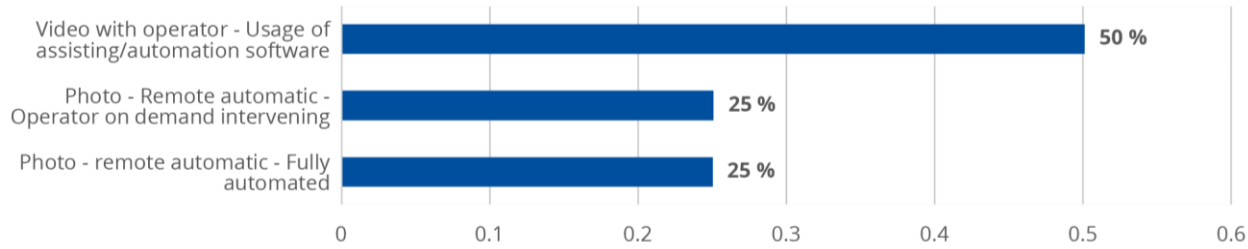


Figure 29: Most effective types of attacks

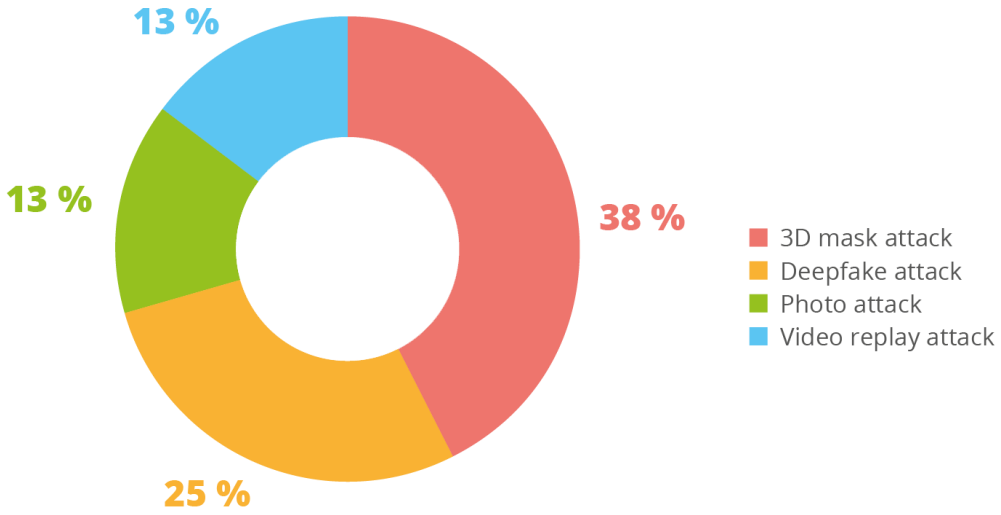
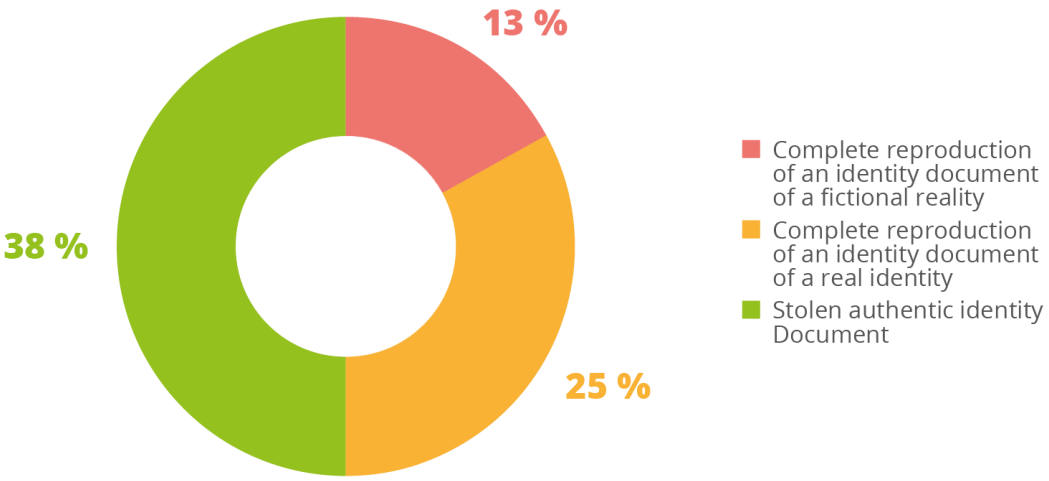


Figure 30: Attack involving ID's most likely to succeed (%)



We then asked for information on the different types of attacks they have experienced themselves. The results clearly show that attacks involving evidence based on an identity document, attacks driven with a stolen authentic identity document are more likely to succeed, according to 38% of respondents. A possible countermeasure to this specific case has been identified in the process of crosschecking the ID of the person who is enrolling with several available data bases of stolen documents.

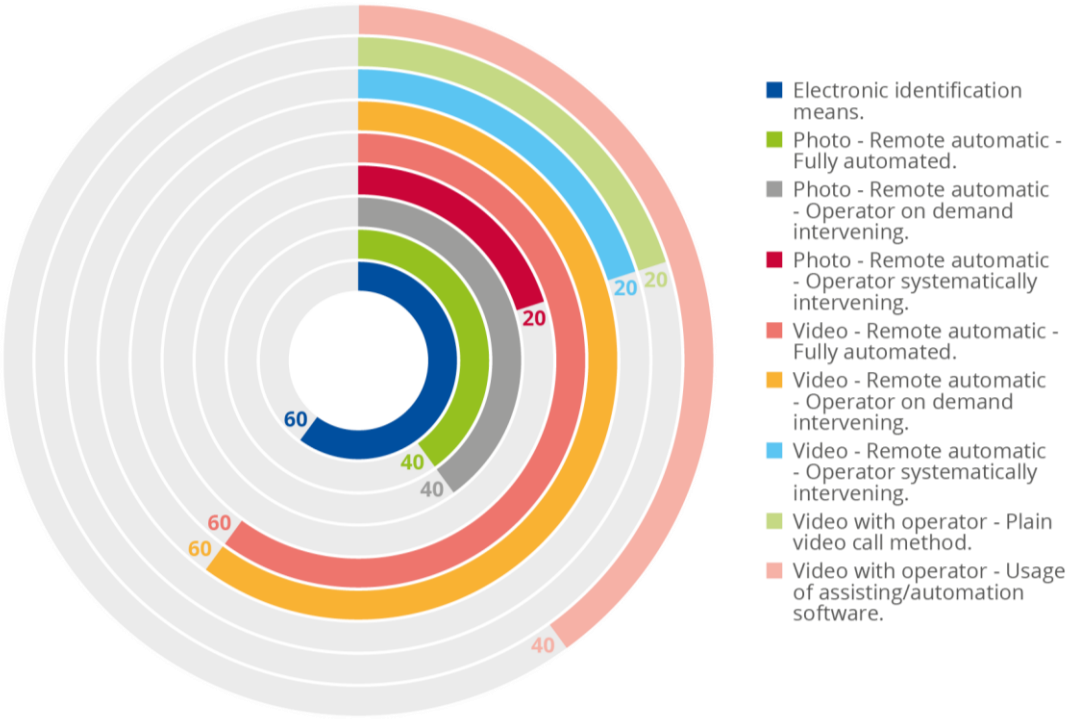
Furthermore, speaking more generally about attacks '3D Mask attacks' (38%) together with 'Deepfake attacks' (25%) received the higher number of votes from respondents who identified them as the most effective types of attacks.

In addition, **the face presentation attack most concerning for the future is the deepfake attack** with 62% of the votes.

B.3 SURVEY RESULTS FROM THE RESEARCHER CATEGORY

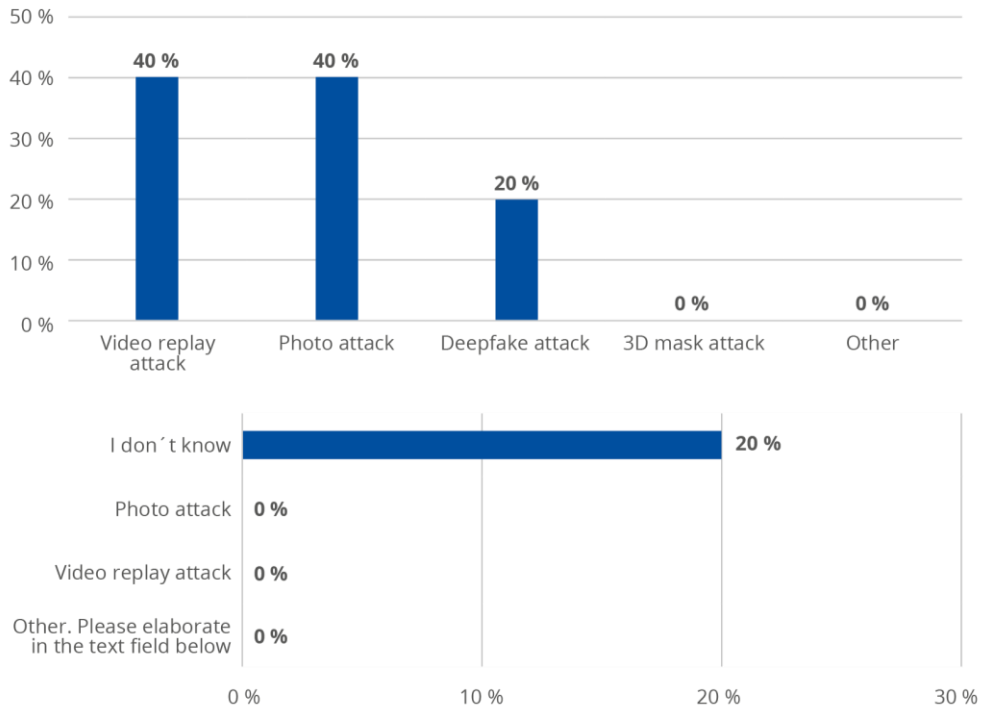
The survey results in the researcher category enabled the research team to understand in which direction R&D is moving in this specific field. Nowadays, with the rivalry of disciplines such as AI and machine learning, researchers are investigating how to make remote identity proofing processes as autonomous as possible. The graph below gives us a picture of the state-of-the-art.

Figure 31: Methods Currently Studied (%)



Regarding the current effectiveness of different types of attacks used in the wild, video replay attacks and photo attacks (low technology and low-cost attack typologies) are the most frequent and likely to succeed. On the other hand, we see deepfake attacks as a repeated pattern across all categories interviewed in terms of spoof effectiveness.

Figure 32: Current Attack Success Ratios



Also, not surprisingly, in this category with a share of 60% of overall respondents, it is believed that the most concerning attacks for the future are indeed deepfake attacks since making them is becoming much easier and the technologies needed are available in the market at affordable prices.

Lastly, focusing on possible countermeasures, respondents as a community agree that a well-defined risk analysis and a well-defined system architecture would definitely ensure a reliable level of security. When the threats are well identified, the right countermeasures can and should be put in place. Systems should be attacked by an evaluation laboratory to better understand the threats.

Continuously monitoring the quality and rates of fraud detection for the methods of identification demonstrated that there are two organisational controls which are important:

- having a feedback process from the relying party so that they have to reliably inform the identity provider of undetected fraud attempts;
- having a 'mystery attacker' process where the identity provider itself is trying to break the process to detect any security vulnerabilities.

C ANNEX: WORKSHOP RESULTS

During the first panel, panellists were asked to answer the following questions:

- Do you see these attacks in the wild?
- How do you fight them?
- Is a human operator still necessary?
- How to increase the cost of frauds?
- Should data intake be photos or videos?

Regarding the need for human operators, identity providers usually ask users to provide a video which is analysed asynchronously. This means that the video is checked both by human operators and ML algorithms. About 6% of identities are refused because they are fraudulent (the number seems small but it is not). About 80% of attacks target identity documents. Many identity providers use and relay in ML and are focused on the detection of liveness. However, most algorithms are focused on biometric data and not on documents. Since there are several types of identity documents, there are many possible threats.

Speaking of countermeasures, panellist explained that both active and passive solutions are used, which however always involve the use of video and operator intervention. The video helps a greater number of security checks to be performed and the operator helps the artificial intelligence to identify any new types of attacks.

Moreover, active solutions are easier to implement but both are required to understand some attacks. Active approaches are easy to use but passive ones can help balance false positive and fraud detections.

Furthermore, if you want to have continuous training to follow the trends of new frauds you need to regularly have new data to train the algorithm and test it. You need humans to tag data, otherwise you can't get successful training. This means that the human operator is essential in the cycle. Algorithms cannot understand and detect new fraud on their own; human action is required to assign the correct labels to new attacks.

Regarding the different types of evidence involved in the process, panellists argued that they prefer video to photos as it provides more data for the verification of identity and can be used to apply more functionality. For example, videos can help detect holograms and other security features that cannot be detected in a single frame. Photos are still used, but they think it's an out-dated approach.

Nowadays, **most remote ID attacks are low tech, with attackers presenting fake IDs or presenting someone else's face on a display (so-called replay attack). However, deepfake attacks are expected to become more frequent and harder to detect.**

Therefore, countermeasures will need to evolve as well. Both active (i.e. asking the user to read a random set of numbers) and passive (i.e. face texture analysis) security controls will have a role in the future. Synergies between AI and human operators will need to be further developed to spot the fakes.

In the future, they think human operators will still have an important role in the remote identity proofing process, although we will need them less and less over time. For low-security environments, a fully automated process might be achieved in a few years.

Deepfakes are a real threat to the future. They will not only have an impact on the remote identification sector but will also affect the press, telecommunications, etc.

Digital identity will become a kind of protocol on the web, a level of security that will create the necessary trust in the web for us to trust each other as we do in the physical world. Digital identity will involve all use cases, the way we do business on the web and the way we meet for personal relationships.

There is a need for interoperability. Artificial intelligence has to be trained to make it more user-friendly, convenient and more automatic while keeping the human being in the loop. A risk-based approach is the only way to go, threats need to be investigated and updated to be one step ahead of what an attacker might do.

Perhaps, in this field there is the need of an agency such as ENISA to investigate those risks and spread their knowledge by organising events to inform the stakeholders in the ecosystem about the threats they could face and share best practices. There is a need to test solutions and set security goals.

There is a need for:

- an independent organisation
- audit capabilities
- testing capabilities.

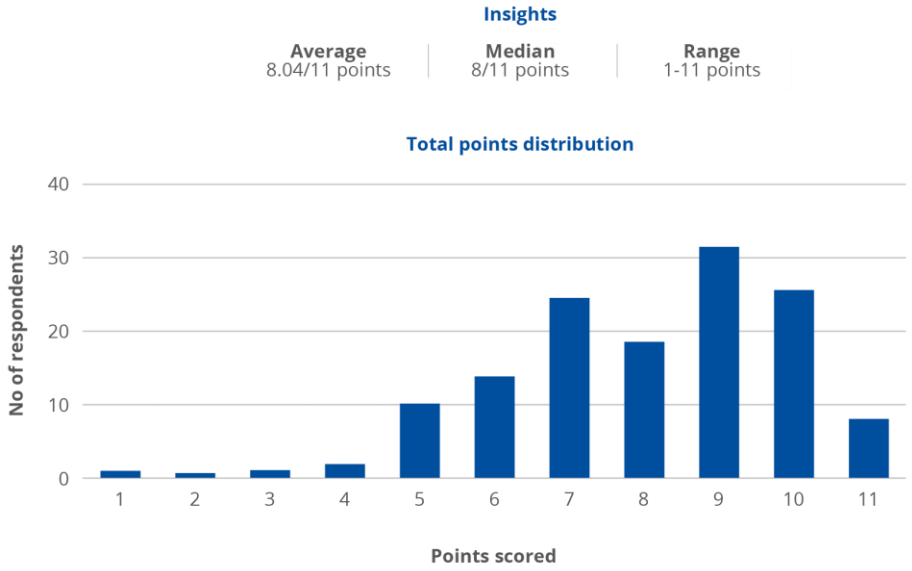
Creating deepfakes in real-time is more complicated and consequently live sessions will become more popular to deal with this type of attack. Audio is also a dimension that must be taken into consideration as a possible countermeasure.

In the future, attackers will seek to exploit inherent vulnerabilities of AI. Serious risk analysis will be necessary because it all depends on the level of risk which is closely related to the abilities and incentives of the attackers. The key is to make the attackers' job more expensive. A key solution could certainly be multifactorial security.

To actively engage participants, several interactive sessions were designed and proposed throughout the conduction of the workshop. The following activities were proposed to the audience:

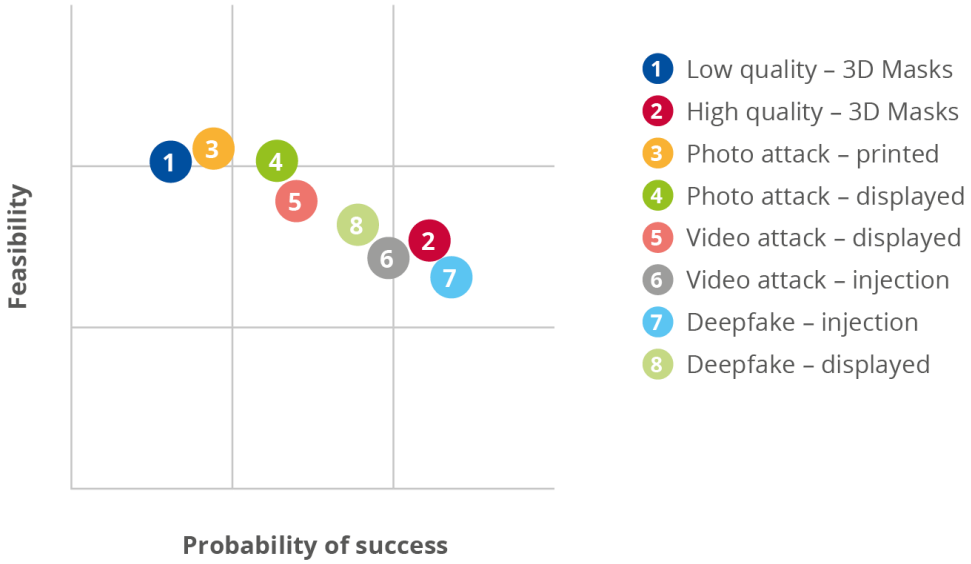
- 3D mask test,
- ranking exercise on the future of spoofs (Mentimeter tool),
- asking participants to identify a deepfake participant hidden among workshop panellists.

Figure 33: 3D mask test results.



For this interactive session we received 143 responses from the participants. The results confirmed us that humans' operators could be fooled in recognising 3D mask presentation attacks. On average 3 mistakes were made from 11 images shown.

Figure 34: Future of spoofs

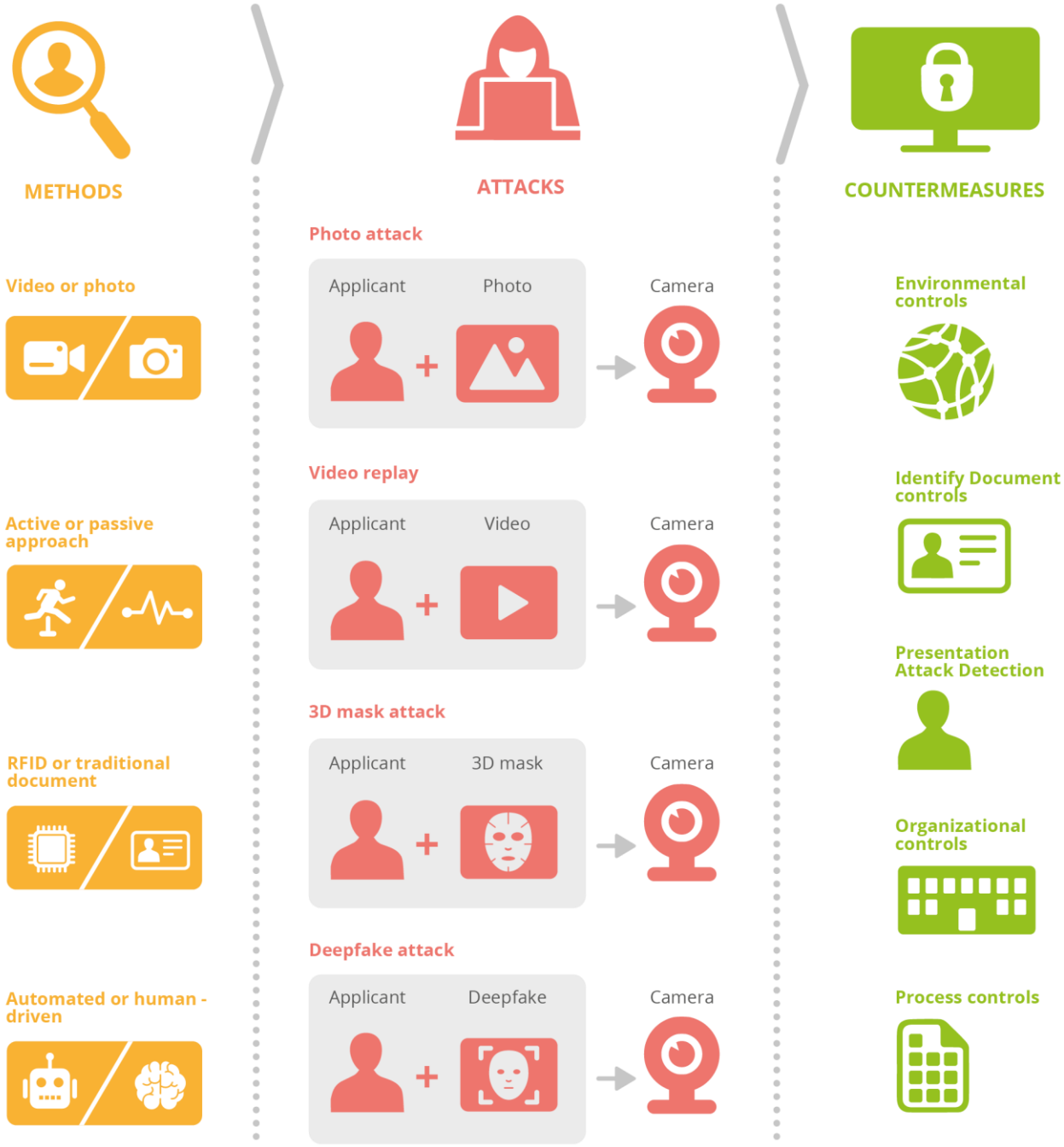


The results showed some homogeneity.

Overall, the results show that the participants are aware of the topic as the matrix shows a very realistic view. From a technical point of view people have a lot of faith in deepfake technology but when experimenting it is realised that a really good quality deepfake still suffers from many other practical problems such as distortion. Deepfake is therefore probably overrated as its chances of success are not very high.

D ANNEX: METHODS, ATTACKS AND COUNTERMEASURES MAP

Figure 35: Map of methods, attacks and countermeasures



TP-09-21-525-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
1 Vasilissis Sofias Str
151 24 Marousi, Attiki, Greece

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

enisa.europa.eu

ISBN: 978-92-9204-549-4
DOI: 10.2824/183066