INTERNATIONAL STANDARD

ISO/IEC 19989-2

First edition
2020-10

Information security — Criteria and methodology for security evaluation of biometric systems —

Part 2: Biometric recognition performance

Sécurité de l'information — Critères et méthodologie pour l'évaluation de la sécurité des systèmes biométriques —

Partie 2: Efficacité de reconnaissance biométrique

Reference number
ISO/IEC 19989-2:2020(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 27, *Information security, cybersecurity and privacy protection*.

A list of all parts in the ISO/IEC 19989 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Biometric systems can be subject to presentation attacks where attackers attempt to subvert the system security policy by presenting their natural biometric characteristics or artefacts holding copied or faked characteristics. Presentation attacks can occur during enrolment or identification/verification events. Techniques designed to detect presentation artefacts are generally different from those to counter attacks where natural characteristics are used. Defence against presentation attacks with natural characteristics typically relies on the ability of a biometric system to discriminate between genuine enrolees and attackers based on the differences between their natural biometric characteristics. This ability is characterized by the biometric recognition performance of the system – how well or badly a biometric recognition system executes its required functions. Biometric recognition performance and presentation attack detection have a bearing on the security of biometric systems. Hence, the evaluation of these aspects of performance from a security viewpoint will become important considerations for the procurement of biometric products and systems.

Biometric products and systems share many of the properties of other IT products and systems which are amenable to security evaluation using the ISO/IEC 15408 series and ISO/IEC 18045 in the standard way. However, biometric systems embody certain functionality that needs specialized evaluation criteria and methodology which is not addressed by the ISO/IEC 15408 series and ISO/IEC 18045. Mainly, these relate to the evaluation of biometric recognition and presentation attack detection. These are the functions addressed in ISO/IEC 19989 (all parts).

ISO/IEC 19792 describes these biometric-specific aspects and specifies principles to be considered during the security evaluation of biometric systems. However, it does not specify the concrete criteria and methodology that are needed for security evaluation based on the ISO/IEC 15408 series.

The ISO/IEC 19989 series provides a bridge between the evaluation principles for biometric products and systems defined in ISO/IEC 19792 and the criteria and methodology requirements for security evaluation based on the ISO/IEC 15408 series. The ISO/IEC 19989 series supplements the ISO/IEC 15408 series and ISO/IEC 18045 by providing extended security functional requirements together with assurance activities related to these requirements. The extensions to the requirements and assurance activities found in the ISO/IEC 15408 series and ISO/IEC 18045 relate to the evaluation of biometric recognition and presentation attack detection which are particular to biometric systems.

ISO/IEC 19989-1 consists of the introduction of the general framework for the security evaluation of biometric systems, including extended security functional components, and supplementary methodology, which is additional evaluation activities for the evaluator. The detailed recommendations are developed for biometric recognition performance aspects in this document and for presentation attack detection aspects in ISO/IEC 19989-3.

This document describes supplements to the evaluation methodology for biometric recognition performance evaluation for the security evaluation of biometric products. It supplements the ISO/IEC 15408 series, ISO/IEC 18045 and ISO/IEC 19989-1. It builds on the general considerations described in ISO/IEC 19792 and the biometric performance testing methodology described in ISO/IEC 19795-1 by providing additional guidance to an evaluator.

In this document the term “data subject” is used while "user" is used in ISO/IEC 19989-1, in order to be consistent with biometric vocabulary, as biometric experts are supposed to be the main readers of this document.

Information security — Criteria and methodology for security evaluation of biometric systems —

Part 2:
Biometric recognition performance

1 Scope

For security evaluation of biometric verification systems and biometric identification systems, this document is dedicated to the security evaluation of biometric recognition performance applying the ISO/IEC 15408 series.

It provides requirements and recommendations to the developer and the evaluator for the supplementary activities on biometric recognition performance specified in ISO/IEC 19989-1.

The evaluation of presentation attack detection techniques is out of the scope of this document except for presentation from impostor attempts under the policy of the intended use following the TOE guidance documentation.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 2382:2015, *Information technology — Vocabulary*

ISO/IEC 2382-37:2017, *Information technology — Vocabulary — Part 37: Biometrics*

ISO/IEC 15408-1:2009, *Information technology — Security techniques — Evaluation criteria for IT security — Part 1: Introduction and general model*

ISO/IEC 15408-3:2008, *Information technology — Security techniques — Evaluation criteria for IT security — Part 3: Security assurance components*

ISO/IEC 18045:2008, *Information technology — Security techniques — Methodology for IT security evaluation*

ISO/IEC 19792:2009, *Information technology — Security techniques — Security evaluation of biometrics*

ISO/IEC 19795-1:2006, *Information technology —Biometric performance testing and reporting —Part 1: Principles and framework*

ISO/IEC 19795-2:2007, *Information technology — Biometric performance testing and reporting — Part 2: Testing methodologies for technology and scenario evaluation*

ISO/IEC 19989-1:2020, *Information security — Criteria and methodology for security evaluation of biometric systems — Part 1: Framework*

ISO/IEC 30107-3:2017, *Information technology — Biometric presentation attack detection — Part 3: Testing and reporting*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 2382-37:2017, ISO/IEC 2382:2015, ISO/IEC 15408-1:2009, ISO/IEC 18045:2008, ISO/IEC 30107-3:2017 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
bona fide presentation
interaction of the biometric capture subject and the biometric data capture subsystem in the fashion intended by the policy of the biometric system

Note 1 to entry: Bona fide is analogous to normal or routine, when referring to a bona fide presentation.

Note 2 to entry: Bona fide presentations can include those in which the user has a low level of training or skill. Bona fide presentations encompass the totality of good-faith presentations to a biometric data capture subsystem.

Note 3 to entry: The concept "in the fashion intended by the policy of the biometric system" for bona fide is included in the concept "in accordance with the policy of the intended use of the biometric system" used in this document

[SOURCE: ISO/IEC 30107-3: 2017, 3.1.2, modified — Note 3 to entry has been added]

3.2
bona fide presentation classification error rate
BPCER
proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario

[SOURCE: ISO/IEC 30107-3: 2017, 3.2.2]

3.3
detection error trade-off curve
DET curve
modified ROC curve which plots error rates on both axes (false positives on the x-axis and false negatives on the y-axis)

Note 1 to entry: An example set of DET curves is shown in ISO/IEC 19795-1:2006, Figure 3.

[SOURCE: ISO/IEC 19795-1:2006, 4.7.1]

3.4
false accept rate
FAR
proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed

[SOURCE: ISO/IEC 19795-1:2006, 4.6.6]

3.5
false-negative identification-error rate
FNIR
proportion of identification transactions by users enrolled in the system in which the user's correct identifier is not among those returned

[SOURCE: ISO/IEC 19795-1:2006, 4.6.8]

3.6
false-positive identification-error rate
FPIR
proportion of identification transactions by users not enrolled in the system, where an identifier is returned

[SOURCE: ISO/IEC 19795-1:2006, 4.6.9]

3.7
false reject rate
FRR
proportion of verification transactions with truthful claims of identity that are incorrectly denied

[SOURCE: ISO/IEC 19795-1:2006, 4.6.5]

3.8
impostor attack presentation match rate
IAPMR
<full-system evaluation of a verification system> proportion of impostor attack presentations using the same PAI species in which the target reference is matched

[SOURCE: ISO/IEC 30107-3: 2017, 3.2.6]

3.9
operating point
setting of a biometric system to operate at a fixed decision threshold

Note 1 to entry: An operating point can be directly represented by a decision threshold or can be represented by a predefined configuration parameter.

3.10
policy of the intended use
policy stating how bona fide presentations are to be made

Note 1 to entry: Intended use is about how natural biometrics should be used with the TOE, i.e. presentations made in the way that bona-fide presentations are made. Presentations made with artefacts are not considered.

4 Abbreviated terms

ATE	assurance class tests
AVA	assurance class vulnerability assessment
FAR	false accept rate
FMR	false match rate
FNIR	false-negative identification-error rate
FNMR	false non-match rate
FPIR	false-positive identification-error rate
FRR	false reject rate
FTAR	failure to acquire rate
FTER	failure to enrol rate
IT	information technology

PAD	presentation attack detection
PAI	presentation attack instrument
PP	protection profile
ST	security target
TOE	target of evaluation
TSF	TOE security functionality

5 Supplementary activities to ISO/IEC 18045 on ATE tests

5.1 General

5.1.1 Guidance

Clause 5 contains guidance, additional requirements and supplements to evaluation activities from ISO/IEC 19989-1:2020, Clause 14, for the evaluator.

The definition of authentication can be found in ISO/IEC 2382.

The definitions of biometric (adjective), biometric capture, biometric capture device, biometric characteristic, biometric enrolee, biometric enrolment, biometric enrolment database, biometric feature, biometric identification, biometric impostor, biometric presentation, biometric recognition, biometric reference, biometric sample, biometric system, biometric verification, comparison, enrol, failure-to-acquire rate, failure-to-enrol rate, false match rate, false non-match rate, match and threshold can be found in ISO/IEC 2382-37.

NOTE 1 In this document, the expression "capture device" is sometimes used instead of “biometric capture device”.

NOTE 2 In this document, the short expression “enrolee” is used instead of “biometric enrolee”.

NOTE 3 In this document, the short expression “enrolment” is used instead of “biometric enrolment”.

NOTE 4 In this document, the short expression “feature” is often used instead of “biometric feature”.

NOTE 5 In this document, the expression "impostor" is sometimes used instead of “biometric impostor”.

NOTE 6 In this document, the short expression “presentation” is often used instead of “biometric presentation”.

The definitions of assurance, attack potential, class, component, confirm, describe, determine, developer, development, ensure, evaluation, guidance documentation, identity, interaction, interface, object, operation, operational environment, potential vulnerability, protection profile, security target, target of evaluation, TOE security functionality, verify and vulnerability can be found in ISO/IEC 15408-1.

NOTE 7 The term "operation" is related to the AGD class.

The definitions of check, examine, methodology and report can be found in ISO/IEC 18045.

The definitions of presentation attack, presentation attack detection and presentation attack instrument can be found in ISO/IEC 30107-1.

Biometric systems employ technology and functionality that require special considerations when conducting security evaluation, including security evaluation based on the ISO/IEC 15408 series. One of these is the non-deterministic nature of biometric decisions, i.e. match; non-match; and other decisions, and the consequent possibility of decision errors (e.g. false match, false non-match) which can have security implications for biometric systems.

A test of the security relevant biometric recognition error rates is an important aspect of every security evaluation of a biometric system. Further, the requirements in ISO/IEC 19989-1 ensure that also the developer of the biometric system under evaluation shall test the error rates of the system under the policy of the intended use following the TOE guidance document.

NOTE 2 In this document, the intended use following the TOE guidance document covers both genuine and imposter attempts, as long as the usage is consistent with the guidance. The guidance is provided by the TOE developer.

This clause contains guidance and additional requirements for the evaluator and review of the developer tests as well as for planning, conducting and reporting independent testing of the error rates of the biometric system. This clause may also be used by the developer of a biometric system to be informed about the requirements.

NOTE 3 In this document, the evaluator is considered competent under the framework of evaluations under ISO/IEC 15408 (all parts) and in particular ISO/IEC 15408-1.

Subclauses 5.1 to 5.4 are applicable to both ATE_IND and ATE_FUN. Subclause 5.5 is specific to ATE_FUN, related to functional tests. Subclause 5.6 introduces the specific aspects resulting from ISO/IEC 19989-1. Subclause 5.7 is for ATE_IND.2 part related to developer tests. Subclause 5.8 is specific to independent testing.

The evaluator should, as a default principle, follow the recommendations introduced thereafter (e.g. on error rates, maximum values, developer testing methodology, etc.). If the evaluator judges they are not appropriately chosen with respect to the TOE and the application, they may use more appropriate values for testing and shall provide justification for the choice of values in the evaluation report. The technology specific aspects in this clause have been developed under consideration of the requirements in ISO/IEC 19795-1. The type of testing to be performed (scenario, technology or operational testing) shall be determined by the evaluator, based on the nature of the TOE and the TOE security target (see 5.1.3 for further information).

In addition to the requirements and recommendations provided in this clause, the evaluator shall also follow the requirements for the assurance components selected by the TOE for the ATE class in ISO/IEC 15408-3 and shall follow the requirements of the corresponding activities in ISO/IEC 18045.

The configuration of the TOE can have an effect on the biometric recognition performance. Hence, the evaluator shall ensure that the TOE configuration for testing complies with the requirements specified in the ST or PP. In particular, when the TOE includes PAD functionality, the evaluator shall check that the PAD functionality is enabled and correctly configured while conducting the biometric recognition performance testing. If both biometric recognition performance and PAD are evaluated for the TOE, the bona fide presentation classification error rate (BPCER) as defined in ISO/IEC 30107-3 should be calculated in biometric recognition performance testing, by additionally recording the output of the PAD subsystem as a supplementary information in the documentation for the ATE_FUN activity of the PAD evaluation. Similarly, the impostor attack presentation match rate (IAPMR) as defined in ISO/IEC 30107-3 may be retrieved from the ATE_FUN activity of the PAD evaluation and taken in account as it is related to biometric performance.

NOTE 4 The information on IAPMR from PAD evaluation is not useful for the evaluator for estimating FMR/FAR, as those metrics are not directly related. Nevertheless, IAPMR can be a useful information for the evaluator to understand specific behaviours of the recognition algorithms (and it can also be useful for AVA to identify potential weaknesses).

NOTE 5 ATE is focused on validating the performance of the TOE by testing under the policy of the intended use following the TOE guidance document. It therefore encompasses bona fide presentation attempts (as opposed to presentation attacks considered in ISO/IEC 19989-3) for both mated-comparison trials and non-mated comparison trials (i.e. imposter attempts). In both trials, the evaluator can assume use of the TOE in accordance with the policy of the intended use. All other kinds of presentations are considered in ISO/IEC 19989-3.

5.1.2 Remarks for performance evaluation

The relationship between the two error rates FAR/FRR can be illustrated using a detection error trade-off (DET) curve showing the dependency between the two biometric error rates as the decision threshold is varied over its working range (see ISO/IEC 19795-1 for more information). DET curves can be useful to compare the recognition performance of biometric systems and to track improvements in a biometric system over its development. In the context of the evaluation, the biometric system is instead usually considered at only one or a very limited set of decision thresholds.

The developer shall specify the operating point or points of the TOE in the security target, in order to ensure that customer of the TOE is informed about the evaluated configuration.

The evaluator shall ensure that the relevant security settings (including at least the operating points) of the TOE used during performance testing are set, and the evaluation performed in accordance with the values stated in the security target.

Further, it should be considered that a security evaluation following the ISO/IEC 15408 series is focused on IT-security. Therefore, the security relevant biometric recognition error rates of a biometric system shall be assessed in the context of an evaluation. However, because security can be achieved at the expense of usability, usability-related error rates should also be evaluated. Guidance on identifying relevant error rates is given in 5.1.4.3.

Subclauses 5.1.3 to 5.8 provide more detailed information for the evaluator regarding the review of developer tests, repeating a test subset as outlined in ATE_IND.2 and regarding the independent test as required by ATE_IND.1.

5.1.3 Identification of the type of performance evaluation

According to ISO/IEC 19795-1, three basic types of evaluation for the performance rates of a biometric system can be distinguished:

— technology evaluation: off-line evaluation of one or more algorithms for the same biometric modality using a pre-existing or specially collected corpus of samples;

— scenario evaluation: evaluation in which the end-to-end system performance is determined in a prototype or simulated application using live biometric presentations made by a test crew recruited for the test;

— operational evaluation: evaluation in which the performance of a complete biometric system is determined in its operational environment with a specific target population.

The first step for the evaluator is to identify the correct type of the evaluation for the biometric system under evaluation. The type of evaluation shall be determined by the composition of the TOE and what is specified in the security target. The composition of the TOE shall be capable of supporting the specified type(s) of evaluation.

NOTE As evaluations usually refer to an instance of a biometric product rather than to a concrete instance of an installation of a biometric system, the operational test of the security relevant biometric recognition error rates is often not considered. Consequently, most of the content in this clause refers specifically only to technology and scenario cases.

The type of evaluation to be performed depends on the definition of the TOE and the scope in the security target. The evaluator shall verify that the developer testing is appropriate to the type of evaluation.

ISO/IEC 19795-1 further distinguishes between online and offline tests. In online tests, the enrolment or comparison process is executed at the time of image or signal submission while those phases of testing are kept separately in offline tests. Technology evaluations are carried out using offline processing of biometric data. Due to requirements regarding the repeatability and reproducibility that apply to evaluations, pure online tests (in which the images or signals are directly discarded) should not be used.

5.1.4 Biometric recognition error rates

5.1.4.1 Metrics for biometric verification

The class ATE refers in this document to the tests that shall be performed to assess the performance of the biometric system under the policy of the intended use following the TOE guidance document. In the case of a verification biometric system the intended use can be defined as follows: "a data subject tries to be recognized by the system as a legitimate enrolled data subject related to a claimed identity".

In this scenario, the system may anticipate two cases that shall be distinguished: biometric mated comparison trial (i.e. as genuine) and non-mated comparison trial (i.e. as impostor). According to these two cases the following decision error rates shall be reported:

— for an algorithm evaluation, the FMR and FNMR;

— for a system evaluation, the FAR and FRR.

The difference between algorithm error rates (FMR and FNMR) and the system error rates (FAR and FRR) is that the latter depends on the permitted number of verification attempts and may also include other type of errors such as the failure to acquire and failure to enrol.

The FAR (respectively FMR) and FRR (respectively FNMR) error rates of a biometric system are inversely related, the trade-off between the two being determined by the verification decision threshold setting for the system.

Note that other error rate testing that includes the sample acquisition stage typically produces different results for transactions that are limited to a single attempt and those that allow multiple attempts, for example: failure to enrol rate (FTER) and failure to acquire rate (FTAR).

5.1.4.2 Metrics for biometric identification

In an identification scenario, a subject provides a biometric sample without making an explicit claim of identity. The biometric system identifies the subject by biometric comparison of the biometric identification sample with the biometric references of all enrolled subjects until a match is found (or not) based on identification decision criteria defined for the system. This is known as 1:many comparison. Depending on the criteria, zero or more matches can be found and reported by the system. When more than one match is reported, the matching identities can be ranked according to the corresponding comparison scores.

In the case of an identification biometric system, the intended use may be defined as follows:

— positive identification scenario: scenario where the purpose of a biometric system is to verify and identify by means of biometric recognition that a data subject is a specific enrolee in the system without requiring a prior claim of identity;

— negative identification scenario: scenario where the purpose of a biometric system is to confirm by means of biometric recognition that an enrolment data subject is not enrolled in the system.

As with the verification scenario, the class ATE activity refers in this document to the tests that shall be performed to assess the performance of the biometric identification system (TOE) under its intended use. Performance testing shall include:

— if positive identification scenario is considered: performance testing for the case of bona fide presentations, i.e. where members of a test crew (or test data) comprising legitimate enrolled data subjects attempt to be identified by the system as themselves, and performance testing for the case of impostor presentations where members of a test crew (or test data) who are not enrolled in the system attempt to be falsely identified by the system as legitimate enrolees using presentations of their natural biometric characteristics;

— if negative identification scenario is considered: performance testing for the case of non-enrollees-related presentations where members of a test crew (or test data) who are not enrolled in the

system attempt not to be identified by the system as enrollees, and performance testing for the case of enrollees-related presentations, i.e. where members of a test crew (or test data) comprising enrolled data subjects attempt to be falsely not identified by the system, using presentations of their natural biometric characteristics.

Contingent on the system functions, the main metrics that should be assessed are:

— if positive identification scenario is considered, true-positive identification rate, false-positive identification-error rate (FPIR) and identification rank;

— if negative identification scenario is considered, true-negative identification rate and false-negative identification error rate (FNIR).

NOTE For any scenario, further biometric system error rates exist that can become relevant for specific scenarios. ISO/IEC 19795-1 gives a complete overview over all possibly relevant biometric recognition error rates.

From these types of error rates, the evaluator shall decide which are relevant for a particular evaluation (see 5.1.4.3).

5.1.4.3 Identification of relevant error rates

There is no comprehensive and single answer to the question of which error rates are relevant for a particular biometric system under a specific evaluation. The evaluator should consider a number of aspects when determining relevant error rates as described below.

This subclause provides an overview of the most important aspects to be taken into account in order to answer this question. The evaluator shall also take into account the claimed error rates (following ISO/IEC 19989-1:2020, 8.3, 8.4 and Annex C).

The primary error rates of interest are those that are security relevant. The error rates that are security relevant depend on the purpose of the biometric recognition. For verification, the primary security metric for scenario testing is FAR (FMR for technology testing). For positive identification, the primary security metric for scenario testing is FPIR (FMR for technology testing). Other parameters can influence the security relevant biometric recognition error rates, like the use of a retry counter, for instance.

If both biometric recognition performance and PAD are evaluated for the TOE, BPCER is an additional relevant error rate. It relates to the performance and usability of a system as it is a measure of the rate at which a PAD subsystem determines that a bona fide presentation is a presentation attack when it is not. In addition to the requirements from ISO/IEC 19989-3, the evaluator should observe the output of the PAD subsystem during the performance testing to measure the false classifications of bona fide presentation as presentation attack. Similarly, the evaluator may consider IAPMR as a relevant metric if provided from the ATE_FUN activity of the PAD evaluation, as it relates to simultaneously succeed to pass PAD and biometric recognition.

During an evaluation, the security relevant biometric recognition error rates shall be assessed in depth. The other related error rates should be assessed.

In order to identify all relevant error rates, the evaluator shall consider all error rates that are defined in ISO/IEC 19795-1 and answer two questions for each rate:

— is the error rate security relevant for the TOE?

— what is their relevance to the application?

Only if both questions have been positively answered, the error rate should be taken into account for the evaluation.

It should be noted that some of the error rates of biometric systems depend on other error rates via the setting for the decision threshold of the system. An example is the inverse relationship between False accept rate and false reject rate. Thus, all error rates that are correlated to the security relevant

biometric recognition error rates shall also be reported. This requirement aims to ensure that usability of the TOE can also be estimated from the evaluation report.

The evaluator shall check that the result of this analysis is consistent with the information that the developer provided in the security target.

Table 1 introduces a synthesis of main errors and their impact on security and other functionality of the TOE, in the case of scenario testing.

Table 1 — Effect of biometric errors on application functions

Application error metric	Application function and impact		
	Enrolment	Verification	Identification
FTER	— Inability to enrol data subjects — Provision of exception handling procedures	Provision and security of exception handling procedures for data subjects who cannot be enrolled	Provision and security of exception handling procedures for data subjects who cannot be enrolled
FAR	N/A	Impostor can be wrongly verified as legitimate data subject	N/A
BPCER (if PAD functionality)	Inability to enrol data subjects	Legitimate data subject can fail to pass the PAD check and thus to be verified	Legitimate data subject can fail to pass the PAD check and thus be wrongly not identified in candidate list
FRR	N/A	Legitimate data subject can fail to be verified	N/A
FPIR	N/A	N/A	Impostor can be wrongly identified as a legitimate data subject candidate
FNIR	Inability to detect a multiple enrolment or enrolment attempt from an existing enrolee	N/A	Legitimate data subject can fail to be identified in candidate list

An example of identifying relevant error rates for a specific use case is discussed in Annex B.

5.1.4.4 Determining maximum values for error rates

The maximum permissible value for an error rate is influenced by factors mostly determined by the application. This subclause provides guidance to the evaluator for determining which values are consistent with the TOE (see also Annex B for a typical example). The evaluator shall also take into account the claimed error rates (following ISO/IEC 19989-1:2020, 8.3, 8.4 and Annex C). In order to ensure consistency with the application context, Table 2 taken from ISO/IEC 29115:2013 may be used by the evaluator to identify the required level of assurance on the basis of an assessment of the severity of the possible impact of authentication errors. Determination of what constitutes minimum, moderate, substantial, and high risk depends on the risk criteria for each of the possible consequences.

Table 2 — Potential impact of authentication errors at each level of assurance

Possible impact	Level of assurance			
	Low	medium	high	very high
Inconvenience, distress, or damage to standing or reputation	minimum	moderate	substantial	high
Financial loss or agency liability	minimum	moderate	substantial	high
Harm to the organization, its programs or public interest	N/A	minimum	moderate	high

Table 2 *(continued)*

Possible impact	Level of assurance			
	Low	medium	high	very high
Unauthorized release of sensitive information	N/A	moderate	substantial	high
Personal safety	N/A	N/A	minimum to moderate	substantial to high
Civil or criminal violations	N/A	minimum	substantial	high

In addition, for the case of a biometric verification system used as the sole authentication factor, ISO/IEC TR 29156:2015, 6.4, suggests to select the following FAR values, that can be used by the evaluator in case of scenario testing for setting the FAR or for deriving other error rates:

— a FAR of less than 0,000 1 % for high assurance;

— a FAR of less than 0,01 % for medium assurance;

— a FAR of less than 1 % for basic assurance.

If the biometric verification system is used in conjunction with other (e.g. knowledge-based) authentication factors, higher FAR values may be acceptable.

If no specific level has been defined, the evaluator shall at least follow existing guidelines, such as FRONTEX guidelines[9]. For instance, the evaluator should evaluate the system at a threshold that corresponds to a security level in terms of the false accept rate (FAR) of at most 0,001 (0,1 %). At this configuration, the FRR should not be higher than 0,05 (5 %).

If the ST claims conformance to a PP that defines the maximum values for error rates, in order to satisfy the evaluation requirements, the measured TOE's error rates shall be below the maximum values.

An example of determining maximum values for a specific use case in discussed in Annex B.

5.2 Planning the evaluation

5.2.1 Overview

Planning for biometric recognition performance testing should include two important considerations. Firstly, to ensure that the test design models as closely as possible the real-world scenario of the intended use of the TOE. Secondly, to ensure that the test results are statistically significant in the context of the intended use of the TOE.

Such statistical significance is basically defined by the test data on which tests are run and on the actual access attempts that are performed with the test data available. Results should be reported according to a mean error rate value and a confidence interval.

This subclause encompasses the recommendations for evaluator to plan and execute testing. If developer testing is considered adequate and valid by the evaluator, the evalutor may limit themself to simpler and reduced testing activities. This choice shall be justified in the evaluation report.

A comprehensive test plan shall be devised and documented. The test plan shall satisfy the following objectives.

— The test setup shall match the intended operation of the biometric system as closely as possible.

— The test plan shall exactly identify the relevant steps to be taken during testing. Specifically, when a deep interaction with a test crew is required (e.g. in a scenario test) the test plan shall clearly describe the flow of the test.

— The test plan shall include a very detailed description of the test data that the evaluation will be performed with. This includes for instance: number of data subjects, number of samples per

data subject, number of acquisition sessions involved, environment and external conditions of the acquisition (e.g. background, illumination, pose).

— The test plan shall include a very clear protocol on how the test data is used or how test data subjects should interact with the TOE in the capture process.

— The test plan shall include whether collected data is needed for re-training of the system. If that is the case, then the collected data should be separated in training and test data and the purpose of each dataset shall be described

— The test dataset shall contain several biometric samples per data subject/biometric instance (e.g. from each finger) and samples from different data subjects to generate both mated comparison trials and non-mated comparison trials

— The test plan shall specify a pre-verification phase, including capture (if applicable) and feature extraction to establish a list of samples to be compared together with the information whether the comparison pair corresponds to a genuine test or an impostor test. For each, the evaluator shall plan to measure the time for the suboperations, the number of errors to acquire (if applicable), and number of errors to extract the samples. The list of samples to be compared shall include both genuine comparison pairs and impostor comparison pairs, and the number of pairs shall be set accordingly to the range of error rates (e.g. FAR and FRR for scenario testing of a verification system) values to be checked by the evaluator.

— The test plan shall specify a comparison phase to compute all comparisons from the list prepared in pre-verification phase. For each comparison, the evaluation plan shall include measure of the time for the operation, the number of errors to compare, store the output score when possible or the decision. Based on the output score and the threshold specified by the TOE (if available), or based on the decision, the number of false accepts and the number of false rejects shall be measured.

— The test plan shall define the metrics to be reported. In addition to the values measured during the operations, the evaluator shall identify the test metrics to be reported and how to derive them from the different errors observed by the evaluator. For scenario testing of a biometric verification system, they shall include FTA (if applicable), FTE, FAR and FRR. If comparison scores are available, they can include computation of DET curve in order to embed a DET curve into the report.

— The test plan shall also include a complete specification of the computed comparison scores and how the performance metrics are deducted in order to support statistical analyis and to ensure full reproductibility.

— The test plan shall include the report preparation which shall report computed values for the different metrics identified and a synthesis of the results.

The test plan shall be designed and the testing shall be conducted in accordance with ISO/IEC 19795-2.

5.2.2 Estimation of test sizes

Acquiring and handling the test data is one of the most challenging and most expensive tasks in each test of a biometric system. According to ISO/IEC 19795-1, the test data shall be as large as practically possible. Note that, in order to obtain statistically reliable results, there is a minimum size of the test data. This depends on various factors:

— the size of the error rate to be measured (the smaller the error rate, the larger the necessary test data numbers);

— the required confidence interval.

NOTE The size of the test data set is not the only factor impacting the level of confidence on the error rate to be measured. The dependencies between the different comparison pairs also impact the level of confidence. Full cross-comparison is possible, but then no more full independencies is granted and the impact on level of confidence would be taken into account.

As the required size of the test data also depends on the results of the test themselves, the required test size is usually only estimated roughly in the context of test planning.

The purpose of independent testing at ATE_IND.2 is to seek to validate, even if with limited confidence, the developer claimed error rates. This can be done with a test size much smaller than the size of developer's test data.

The number of test data subjects and test transactions needed to ensure statistical significance of a measured security relevant biometric recognition error rate should be determined from the permitted maximum value of the error rate using either the "rule of 3" or the "rule of 30" (ISO/IEC 19795-1).

For biometrics on mobile devices, practical guidance to conduct the performance testing is provided in ISO/IEC 21879. The evaluator should refer such document to estimate the appropriate test sizes.

5.2.3 Test documentation

It is essential to plan the complete documentation for the test before starting any other activities.

As introduced in 5.1.1, the evaluator should, as a default principle, follow the recommendations introduced thereafter (e.g. on error rates, maximum values, developer testing methodology, etc.). The evaluator may decide from a consideration of the developer test plan and the conduct of developer testing, that the developer testing is adequate for the purpose of the evaluation and the evaluator does not need to repeat entirely the developer testing. Otherwise, the evaluator should follow his/her own test plan and the documentation shall follow the same principles as for any other test in the context of security evaluation following the ISO/IEC 15408 series and ISO/IEC 18045. In addition, the following aspects shall be addressed specifically:

— the exact type of performance evaluation should be carefully described;

— the relevant error rates shall be identified and their maximum acceptable values shall be identified and justified;

— special characteristics of the test data shall be documented;

— the size and characteristics of the test data should be described with special attention to number of sessions involved in its acquisition, data subjects and samples per data subject;

— it should be clearly described how are the different parameters of the system trained/set (if there are any);

— it should be clearly described how the different score sets (mated and non-mated comparison trials) are computed.

5.3 Data collection

5.3.1 Choice of test data or acquiring test crew and capture device

This subclause applies to both scenario and technology testing. An important consideration is the source of the biometric sample data and the associated test crew. This can have been acquired in various ways, for example:

1) live capture: the test data are captured from a test crew specifically for the TOE evaluation (which could involve specific capture requirements, e.g. some particular illumination or background setup in the case of a face recognition system);

2) re-use of pre-existing database, either captured by the developer and/or the evaluator on a previous similar biometric system or obtained from acquired data from third parties (such as the multiple public or private biometric databases available today for benchmarking purpose).

In all cases, the ground truth shall be known i.e. which samples in the acquired data are true matches and which are not.

The most desirable case would be option 1), in which a live capture database is acquired for each evaluation. However, this is also the most time- and resource-consuming solution and a final decision should be adopted on a case-by-case basis. For instance, for a technology or a scenario evaluation, the second case can be sufficient if no specific contextual or external features have to be met (e.g. specific acquisition sensor). If the evaluator chooses to reuse an already existing data, he/she should ensure that this dataset is sequestered and was not accessible by the developer. The main disadvantage of reusing already existing data, that the developer can have access to, is that it can be used to tune their system. The evaluator shall mitigate it by using more sequestered data.

For the acquisition of a new database or test crew, several important factors shall be taken into account in order to obtain results as precise as possible. Among such factors, some ideal characteristics that should be met by a biometric evaluation database are highlighted below.

— The choice of the test data (live capture process or re-use of pre-existing data) shall be under the sole control of the evaluator. As the quality of the test data is essential for the results of the tests, it is important that the evaluator has a detailed knowledge of the acquisition process.

— A question that can often arise is whether test data that has been acquired by the developer beforehand can be re-used during an independent evaluation. While the final decision on the re-use of test data is the decision of the evaluator, this guide encourages the re-use of test data within certain limits. Specifically, a test shall never be based completely on test data that has been acquired by the developer beforehand. Instead, the evaluator shall acquire a small subset of test data and replace it in the original set of test data before using it.

— If the biometric system is designed to work with a specific data subject profile (e.g. men, Asian, over 65 years of age, right-handed), the data subjects in the database/crew should be as close as possible to that profile.

— If the biometric system is not designed to work with one specific sensor, it is better to capture the same individuals with different acquisition devices so that the final evaluation is more general. It is also better to be able to obtain interoperability results (i.e. comparison results between enrolled and test biometric samples captured with different devices).

— The data set should be organized to enable an explicit identification of each sample. However, to ensure personal data protection, the identifier of a sample shall not contain any information related to the real identity of the data subject

— A sufficiently large number of individuals should be enrolled in the database in order to obtain statistically significant results. Such numbers depend on the maximum error rates allowed for the system. The lower the error rates, the larger the number of required data subjects in order to achieve reliable results.

— Also, the different samples of the same data subject should not be captured consecutively but leaving enough time between them in order to simulate the intra-subject variability of the biometric traits. Ideally, the database should be acquired in different sessions separated several weeks among them.

— If relevant, other metadata related to the data subjects can be also acquired. This can include for instance gender, age, use of visual aids (e.g. glasses) or handedness. These metadata can help to further tune the performance evaluation or to reuse the biometric data in future evaluations.

— Biometric data acquisition is prone to various errors such as missing samples, invalid samples, low quality samples and ground truth errors where samples are incorrectly assigned to subject IDs. Some of these errors can be the result of human fallibility and can be reduced by the application of automated data acquisition and recording techniques.

— Biometric data is personally identifiable information (PII). As such, national privacy legislations shall be identified during the acquisition process.

NOTE National data protection laws in the evaluator country can exist and can require, during the acquisition process, informing the acquired data subjects of the use that will be made of their data and obtaining from them a signed consent form for the acquisition of those data.

5.3.2 Performing test

Subclauses 5.2.2, 5.2.3 and 5.3.1 define all the initial steps that should be performed and documented in the test plan prior to the evaluation, that is, relevant error rates and their maximum values, type of evaluation, database and evaluation protocol associated to it, acquisition of data. Once all those steps have been covered, the evaluation should be run according to the predesigned plan. During the evaluation, several aspects shall be documented such as:

— any significant deviations from the original test plan;

— time required to perform each experiment considered in the evaluation. Other temporal information, that can be recorded depending on the evaluation, is the system response time for every access attempt.

Once the tests are performed the results should be reported using standard metrics (see 5.1.4.3 and ISO/IEC 19795-1).

5.4 Analyses

The analysis of the results should not be restricted to the determination of the error rates but should also include a more in-depth investigation of the error instances and whether those errors disclose some security problem of the system [for instance, a significant variability in performance for certain data subject profiles (e.g. men vs women)]. This type of analysis can help to identify potential problematic working scenarios for the system.

It should be noted that, strictly speaking, such an analysis would rather belong into the area of the AVA class than the ATE class as it would open the path to a potential vulnerability of the TOE. Potential security issues found during ATE testing should be reported to the AVA vulnerability assessment activity for further investigation.

These possible deviations from the average expected performance of the system should be reported in the final documentation so that it is clear under which circumstances the system behaves as expected (under the maximum allowed error rates) and in which scenarios the error rates can increase.

A test of the performance of a biometric system in the course of a security evaluation focuses on security relevant biometric recognition error rates such as FAR. Other error rates, e.g. FRR, are relevant to other factors such as usability. Error rates that are security relevant (such as the FAR) can sometimes be traded off against a mutually dependant error rate (in this case the FRR) via the recognition decision threshold setting of the system. Consequently, the dependant error rates shall be evaluated and reported at each operating point specified for the evaluation.

5.5 Reviewing developer tests

In the course of the evaluation activities around ATE_FUN.1, the evaluator shall evaluate the test documentation and results that are provided by the developer.

The developer should follow the requirements from ISO/IEC 19795-1 and ISO/IEC 19795-2 for their performance test. Any deviation shall be justified in the test plan.

The full set of information about the test shall be handed over to the evaluator during the evaluation. Only this way can it be ensured that the evaluator gets a complete overview over all details of the test. In this context, the developer shall provide the evaluator with complete access to the test equipment and test data used for developer testing.

The evaluator shall assess the developer test documentation to confirm that:

— the developer test plan, the conduct of the developer testing and the test documentation conform to the requirements stated in ISO/IEC 19795-1 (see Annex C for an example);

— any deviations from the above requirements are justified;

— the test results show that the security relevant biometric recognition error rates are in accordance with the claims in the TOE security target.

5.6 Specific requirements on assurance components on ATE_IND

5.6.1 Overview

As described in ISO/IEC 19989-1, the following elements of assurance components require the evaluator to conduct their own test of the TOE:

— ATE_IND.1.2E: The evaluator shall test a subset of the TSF as appropriate to confirm that the TOE operates as specified;

— ATE_IND.2.3E: The evaluator shall devise and execute a sample of independent tests and document them to assess the developer test results."

All evaluations of biometric TOEs shall include the assurance component ATE_IND.2 as supplemented by ISO/IEC 19989-1.

Subclauses 5.2.2 and 5.2.3 contain dedicated guidance regarding the evaluation of those two components.

5.6.2 Specific requirements on ATE_IND.1

The specific requirements introduced in ISO/IEC 19989-1 are presented in Table 3. The evaluator derives a subset of the TSF to be tested independently in accordance with the guidance in ISO/IEC 18045. As the biometric performance is an essential part of the TOE, the evaluator shall in any case ensure that this part of the TSF falls into the subset.

Table 3 — Specific requirements on ATE_IND.1 from ISO/IEC 19989-1

ATE_IND.1-3	The evaluator ***shall also devise*** independent testing for performance evaluation setting up a test crew or a test dataset.
ATE_IND.1-4	The evaluator ***shall produce*** test documentation for performance evaluation which satisfies the relevant requirements from ISO/IEC 19795. The evaluator ***shall explain*** any deviation from the test procedures specified in ISO/IEC 19795 and ***also shall describe*** any potential effects and implications for the test results in the test documentation.
ATE_IND.1-5	The evaluator ***shall conduct*** testing using test crew which the evaluator arranged or test data which the evaluator possesses.
ATE_IND.1-6	The evaluator ***shall record*** information of test crew or test data as specified in ISO/IEC 19795.
ATE_IND.1-8	The evaluator ***shall also report*** in the ETR the evaluator testing effort on biometric recognition performance in terms of test size, time spent and also dataset characteristics.

The conducted tests shall follow the requirements and recommendations listed in 5.2, 5.3 and 5.4.

5.6.3 Specific requirements on ATE_IND.2

The specific requirements introduced in ISO/IEC 19989-1 are presented in Table 4. Biometric testing is complex, time-consuming and expensive. Specifically, the acquisition of sufficient test data is a challenge for every test. It is therefore an essential question whether the test data that the developer used can be completely re-used when repeating the test from the developer's test documentation in the context of ATE_IND.2. Guidance on this subject is provided in 5.8.

Table 4 — Specific requirements on ATE_IND.2 from ISO/IEC 19989-1

ATE_IND.2-6	The evaluator ***shall also devise*** independent testing for performance evaluation setting up a test crew or a test dataset.
ATE_IND.2-7	The evaluator ***shall produce*** test documentation for performance evaluation which satisfies the relevant requirements from ISO/IEC 19795. The evaluator ***shall explain*** any deviation from the test procedures specified in ISO/IEC 19795 and ***also shall describe*** any potential effects and implications for the test results in the test documentation.
ATE_IND.2-8	The evaluator ***shall conduct*** testing using test crew which the evaluator arranged or test data which the evaluator possesses.
ATE_IND.2-9	The evaluator ***shall record*** information of test crew or test data as specified in ISO/IEC 19795.
ATE_IND.2-11	The evaluator ***shall also report*** in the ETR the evaluator testing effort on biometric recognition performance in terms of test size, time spent and also dataset characteristics.

The conducted tests shall follow the requirements and recommendations listed in 5.2, 5.3, 5.4 and 5.7.

5.7 Assessing developer tests by repeating a test subset

The requirements of ATE_IND.2 mandate the evaluator to repeat a subset of the developer tests. Tests of the security relevant biometric recognition error rates shall form part of the subset to be repeated.

When repeating a test of the security relevant biometric recognition error rates of the biometric system, it often comes to the question whether it is sufficient to simply repeat the test of the developer. This is of specific relevance if the developer managed to separate the test data acquisition from the actual test of the biometric algorithm. In those cases, the evaluator can decide to simply repeat the test.

The possibilities for the evaluator also depend on the type of biometric test followed by the developer:

— in the case of a technology test, the evaluator has a potential choice of using the same biometric data used by the developer or obtaining a new source of biometric data;

— in the case of a scenario test, the evaluator might be able to use the same test crew as the developer. Otherwise, a new test crew would have to be recruited. In case the evaluator judges that the repetition of a subset of developer tests can be restricted to a technology test, he/she may reuse biometric data obtained from the developer test crew (if available) or use a new source of biometric data.

When possible, the evaluator should take benefit of the efforts the developer spent for test data acquisition by re-using this data when executing testing. On this basis, when possible, the evaluator is encouraged to follow this strategy.: re-use the test data of the developer when executing a subset of the developer tests. In order to avoid a pure repetition of the test using exactly the same data, the evaluator shall consider replacing a subset of the test data by their own data (that should be acquired by the evaluator according to 5.3.1). It falls into the responsibility of the evaluator to decide about the size of this subset. They shall consider the overall quality of the test data of the developer and the quality of the acquisition process (based on its documentation). Instead of replacing a subset, the evaluators can supplement the developer data by their own data. Technically, it is essential that the exchanged subset of the data or the supplemental dataset is large enough to ensure that the developer cannot tune their algorithm based on the database.

Biometric data is personally identifiable information (PII). As such, national privacy legislations shall be identified during the testing process, even while repeating developer's test with some existing test data.

5.8 Conducting independent testing

5.8.1 Overview

ATE_IND.1 (as well as ATE_IND.2 and ATE_IND.3) requires the evaluator to conduct their own test of the security relevant biometric recognition error rates of the TOE. Subclause 5.8 provides the evaluator with the corresponding guidance. Figure 1 summarizes the different steps that shall be performed by an evaluator when planning, conducting and reporting independent testing of the security-relevant biometric recognition error rates of a biometric system.

As introduced in 5.1.1, the evaluator should, as a default principle, follow the recommendations introduced in this clause (e.g. on error rates, maximum values, developer testing methodology, etc.). If the evaluator judges they are not appropriately chosen with respect to the TOE and the application, he/she shall justify other choices in the evaluation report.

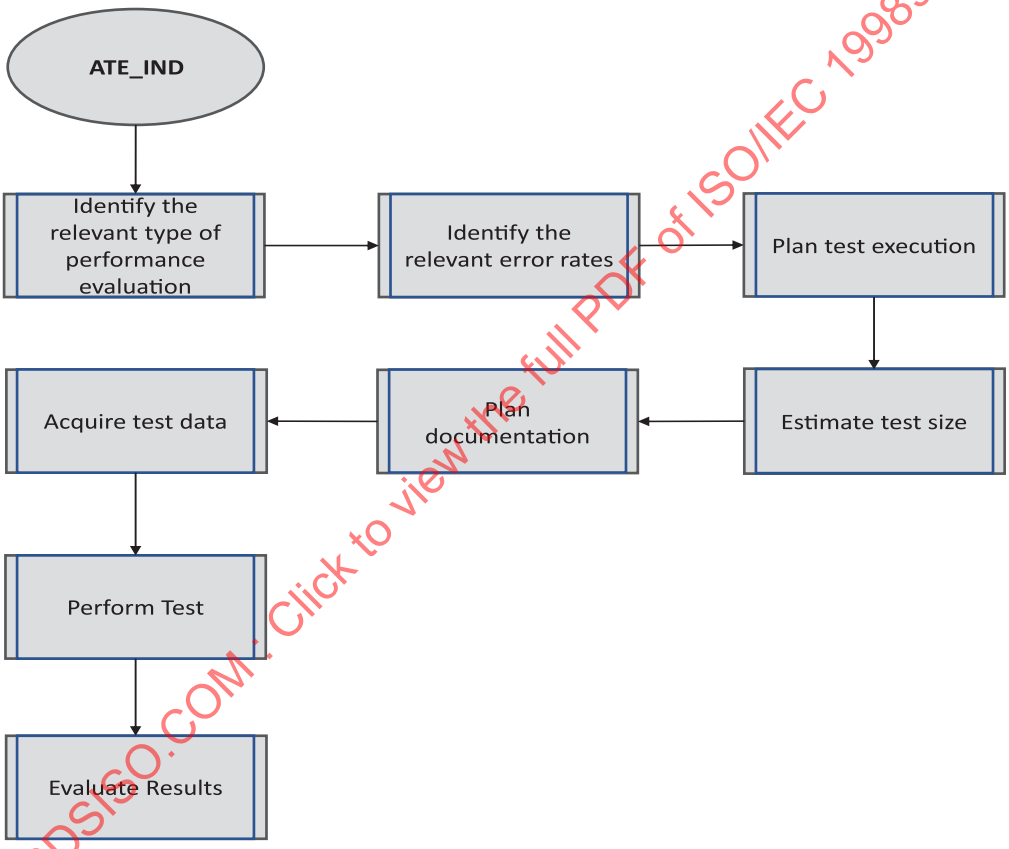

Figure 1 — Processes for independent testing

The process can be divided into the following steps that are described in more detail in subclauses 5.1 to 5.4.

— Identify the relevant type of performance evaluation: Various kinds of test approaches are available starting from a technology performance test of a biometric algorithm to an evaluation of the operational performance test of the biometric system. The correct test approach highly depends on the functionality provided by the TOE.

— Identify the relevant error rates: As a security evaluation focuses on the security-relevant biometric recognition error rates only, not all error rates of the biometric system are relevant. The identification of the relevant error rates is performed based on the type of the biometric system and its application case as defined in the security target.

— Plan test execution: The actual test execution shall be planned in advance.

— Estimate test size: Collecting test data takes a significant amount of the effort of the overall test. It is essential to develop an idea about the amount of test data that is required before starting the actual process of test data acquisition.

— Plan documentation: It is essential to plan the required documentation for the test in advance of the test itself.

— Acquire test data: Test data comprises suitable biometric samples which shall be obtained from live presentations by members of a test crew or using a pre-existing corpus of biometric samples, depending on the type of test (scenario or technology) and the test methodology. In both cases, the test data shall include so called ground truth, that is knowledge of the source of each sample (e.g. the subject ID) such that, with the corpus, both mated-comparison trials and non-mated comparison trials can be executed. To ensure the quality of the test results the evaluator shall utilize test data that is not known to the developer of the TOE.

— Perform test: The test shall be carried out under the sole control and responsibility of the evaluator.

— Evaluate results: Test results shall be evaluated and reported using standard metrics.

The evaluator should follow the requirements from ISO/IEC 19795-1 for their performance tests. Any deviation shall be justified in the test plan and documented in the test report.

In the case of technology testing using a pre-existing corpus of biometric samples, the evaluator shall provide justification that the corpus satisfies the needs of the test and is independent from the developer test data.

5.8.2 Identification of the type of performance evaluation

The type of evaluation to be performed mainly depends on the definition of the TOE and type of the developer's performance evaluation. The evaluator should verify that the developer's type of evaluation is appropriate. If this is the case, the evaluator should follow the same. If not appropriate, the evaluator is free to conduct the independent test following another type. If the same type of evaluation is not followed by evaluator, then the evaluator should justify the motivation in the testing report. If the TOE is made of software functions only, it is likely that the developer will decide to perform a technology evaluation and evaluator should, if appropriate, perform the same technology evaluation. If the TOE is a complete biometric system, in particular with a sensor, it will more likely be tested in a scenario evaluation by the developer and, thus, the evaluator should perform a scenario evaluation. Note that if a TOE does not provide the comparison and decision functions, its biometric recognition performance cannot be tested. It should also be noted that scenario testing can include offline testing of the software algorithms for the purpose of cross-comparison testing for the measurement of performance achievement rates and error rates and for the determination of DET plots.

6 Supplementary activities to ISO/IEC 18045 on vulnerability assessment (AVA)

6.1 General aspects

This clause introduces provisions and supplements to evaluation activities from ISO/IEC 19989-1:2020, Clause 15, regarding vulnerabilities which result from biometric recognition errors that may occur when presenting non-mated natural biometric characteristics or when injecting or altering non-mated biometric characteristics after the data capture subsystem. Vulnerabilities associated with presentation attack detection errors for biometric presentation attacks using PAIs are addressed in ISO/IEC 19989-3.

The purpose of this clause is to introduce the context and provide general guidance to evaluators. Each vulnerability being a specific case, the examples in Annex A are there to provide some concrete guidance to the evaluator. However, the evaluator shall devise his/her own strategy based on his/her expertise and the TOE instance under evaluation.

In addition to the requirements and recommendations provided in Clause 6, the evaluator shall also follow the requirements for the assurance components selected by the TOE for the AVA class in ISO/IEC 15408-3 and shall follow the requirements of the corresponding activities in ISO/IEC 18045.

The objective of the evaluation task is to search for vulnerabilities and to demonstrate that the TOE is resistant to a certain and predefined attack potential.

NOTE 1 If the TOE includes PAD functionality, the evaluator assesses the vulnerabilities in the context of the whole system to take into account the potential interactions between the subcomponents.

NOTE 2 Referring to the threats listed in ISO/IEC 19989-1:2020, 6.1, a) to j), the threats and evaluation activities addressed in ISO/IEC 19989-2 for the vulnerability assessment of biometric verification or identification systems mainly relates to the potential exploitation of a) Performance limitations, h) Hostile environment, i) Procedural vulnerabilities around the enrolment process, for attacking the system, together with the possible impacts of j) Leakage and alteration of biometric data, e) Similarity due to blood relationship, f) Special biometric characteristics, and g) Synthesized wolf biometric samples that can help the evaluator to identify specific weaknesses and to mount more efficient attack.

The approach taken here is a counter-example demonstration: for the evaluation of a TOE at a targeted level of security resistance, i.e. at a given targeted minimum value of attack potential for vulnerabilities found on the TOE, if one attack with an attack potential lower than the targeted level is found, then it is demonstrated that the TOE is not resistant to this level. So, the evaluation focusses on finding one attack applicable to the TOE in its usage context and with a rating lower than the targeted level.

The definition of a successful attack shall be done regarding the security definitions of the security target (ST) and should generate a failure in the security objectives of the TOE (access to forbidden data, unauthorized operation, etc.).

The knowledge used by the evaluator to define an attack and an attack path is all the available knowledge, including a specific background in the area (to be accredited a laboratory has to demonstrate its competence in the area), publicly available knowledge (web, dedicated conferences, etc.) but also its knowledge of the TOE gained from other evaluation tasks (for example, knowledge of the implementation if targeting AVA_VAN.3 or higher).

The objective of an evaluation test is to determine whether the TOE is vulnerable to a specific attack at the targeted attack potential (for instance inspired from the threats listed in NOTE 2). Preparatory to testing the evaluator should perform a preliminary attack potential rating based on the attack specifications (what the attack is, how the attack testing is conducted and how the results are interpreted). The preliminary attack rating may be used to prioritize the testing (the lower the rating, the higher the priority). The evaluator may test attack paths for which the preliminary rating is above the target attack potential.

Once the tests list is established, the evaluator executes the corresponding attacks and, in case of success performs a final rating. This rating is then used to validate or refute the resistance level of the TOE.

NOTE 3 The developer often expects some more information from the evaluation: what are all the weaknesses of my product? What are all the attacks to counter in a certified product? This information is an added value for the evaluation report. However, it is not strictly required by the security evaluation under the ISO/IEC 15408 series and ISO/IEC 18045 where a single successful attack can stop the evaluation process, and there is no guarantee that, in case of successful attack, all the possible attacks have been tested and that the TOE is resistant to all the other possible attacks.

6.2 TOE for testing

The developer shall provide the TOE for testing and the TOE shall be suitable for testing.

The exact definition of the TOE delivered shall be done by reference to the guides: any option, configuration, parametrization referenced in the subjects or administration guide should be available to the evaluator. In particular, a biometric system should allow the evaluator to enrol specific people.

In addition, when extra equipment is required to use the TOE, the developer shall make it available to the evaluator (for example, if the TOE is defined as a biometric capture device/sensor, hardware and/or software for connection to a computer and acquisition, processing and exploitation of data).

In some cases, emulators or simulators exist and are used by the developer to validate part of the TOE (for example, the comparison process and its validation over a large database of images). These should be made available to the evaluator.

6.3 Potential vulnerabilities

The vulnerabilities that an evaluator should analyse shall at least take into account those introduced in ISO/IEC 19989-1:2020, 6.1.

Additionaly the evaluator shall consider the combination of those vulnerabilities with other IT-related vulnerabilities. For instance, when considering the possibility to mount a hill-climbing attack, the evaluator shall not restrict the feedback loop to the score reading but shall also estimate the possibility to deduce information related to the score (e.g. based on execution timing or side-channel leakage).

NOTE As highlighted in ISO/IEC 19792:2009, Clause 6, the evaluator also considers the whole system and interactions between subcomponents, as a vulnerability of a subcomponent can be compensated by another subcomponent.

The evaluator shall follow ISO/IEC 19792:2009, Clause 8, to assess the potential vulnerabilities that may be combined with the biometric recognition performance limitations.

6.4 Rating attack potential

For the AVA (vulnerability assessment) class, the rating (i.e. rating of the different factors, calculation of attack potential and comparison with the targeted level) shall be done according to ISO/IEC 19989-1, as well as the consideration of the attack potential (see ISO/IEC 19989-1:2020, Annex F).

Rating examples related to biometric recognition performances vulnerabilities are provided in Annex A.

Annex A
(informative)

Examples of attack potential computation for AVA activities

A.1 General

This annex provides several examples that include various systems that can be evaluated (access control device for a building, an office, etc., access control to a personal device) and the "classical" attacks that can be applied. The attack potential computations are made according to ISO/IEC 19989-1:2020, Annex F.

A.2 Hill climbing

EXAMPLE Consider a fingerprint-based system operating in an uncontrolled environment (for example, protecting the access to an item of equipment). Assume that there is a way to easily connect a computer just before the comparison process, enabling a program to inject probe samples in a suitable format for comparison and that the comparison score is available (for example, through a debug connection). The attack consists in injecting probe samples (for example, starting with random locations of minutia) and to optimize the location using the comparison score in a so-called "hill climbing" attack.

This scenario of attack can correspond to two different objectives: One can be related to privacy leakage (to learn information on the enrolled data), a second one can be related to forging authentication (i.e. an attacker seeking to achieve authentication against a specific enrolee).

It is considered that:

— elapsed time: The identification phase corresponds to find the right interface to the system (connecting a computer to the targeted signals, enabling the presentation of built templates) and to get or write the optimization software for the template generation. It is considered as easy in this example. The exploitation phase is just running the program to get access. 2 weeks for identification and 1 day for exploitation are realistic;

— expertise: Even if the attack method is known and published, setting up the right connections, exploiting specific signals and adapting an optimization software is considered to require an expert level for identification and a proficient level for exploitation;

— knowledge of the TOE: A deep knowledge of the TOE is required (template formats, internal protocols, etc.) so a sensitive knowledge is required;

— window of opportunity (access to the TOE): The TOE operates in anuncontrolled environment (rated as easy for exploitation phase) and it is considered as easy to buy system (rated as easy for identification phase);

— window of opportunity (access to biometric characteristics): It is rated as immediate as the attack does not require the availability of real data (synthetic templates);

— equipment: A specialized equipment is required (computer, connection to the system, and specialized mostly because of template generation, optimization software) for identification. For exploitation, as the software is available the equipment is rated standard.

With the above assumptions, the summary of the corresponding attack potential calculation is provided in Table A.1.

Table A.1 — Calculation of attack potential for the scenario in A.2

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				16	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
2	0	4	4	4	-	0	0	-	0	2	0	12	4

The rating for the attack is enhanced-basic.

If the attack can be performed successfully and no other successful attack with a lower rating is found, the resistance of the TOE is basic.

The system is compatible with the AVA_VAN.2 component.

A.3 Combined attack (getting comparison scores in an indirect way)

EXAMPLE Assume the same system as the previous example, except that the comparison score signal cannot directly be accessed. However, an indirect observation of the comparison process, for example through power consumption or processing time, can give an information of the comparison score (this method is widely used in smart-card evaluations and is known as side channel attacks).

It is considered that:

— as in the previous example, the identification phase corresponds to find the right interface to the system [acquiring internal signals (consumption, time), connecting a computer to the targeted signals, enabling the presentation of built templates] and to get or write the optimization software for the template generation. The exploitation phase corresponds to implement the signal acquisition and processing to the real TOE and run the optimization program to get access;

— elapsed time: More than 1 month for identification and 1 day (less than a week) for exploitation are realistic;

— expertise: Multiple expertise is required [electronics, signal acquisition and processing, biometrics (template format and generation), optimization] for the identification phase. Even if scripted, an expert level is required for exploitation (software for template generation and optimization is written and has just to be used but physical instrumentation of the TOE has to be done);

— knowledge of the TOE: A deep knowledge of the TOE is required (template formats, internal protocols, etc.). Sensitive knowledge is required;

— window of opportunity (Access to the TOE): The TOE operates in a fully uncontrolled environment (rated as easy for exploitation phase) and it is considered as easy to buy system (rated as easy for identification phase);

— window of opportunity (access to biometric characteristics): It is rated as immediate as the attack does not require the availability of real data (synthetic templates);

— equipment: Multiple specialized equipment is required for the identification phase (signal acquisition and processing, computer, connection to the system, template generation, optimization software). For exploitation, as the software is considered as written, specialized equipment is enough.

With the above assumptions, the summary of the corresponding attack potential calculation is provided in Table A.2.

Table A.2 — Calculation of attack potential for the scenario in A.3

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				36	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
8	0	8	8	4	-	0	0	-	0	4	4	24	12

The rating for the attack is high.

If the attack can be performed successfully and no other successful attack with a lower rating is found, the resistance of the TOE is moderate.

The system is compatible with the AVA_VAN.4 component.

A.4 Inverse biometrics attack

EXAMPLE Assume the system is an iris-based system operating in an uncontrolled environment (for example, protecting the access to a device or an equipment). There is an easy way to connect a computer just before the feature extractor, enabling a program to inject synthetic samples and that there is a signal corresponding to the comparison score (for example, through a debug connection). The attack consists of injecting reconstructed synthetic samples and optimizing them using the comparison score.

Note the main difference between this inverse-biometrics attack and the hill-climbing attack described in A.2 is the entry point of the attack, which is before the comparator for hill-climbing (i.e. after the feature extractor) and before the feature extractor for the inverse biometrics attack. Therefore, the inverse-biometrics attack requires more expertise in identification, as a realistic input to the feature extractor needs to be generated in each iteration of the attack (this is why the attack is called inverse biometrics) so the resulting rating is higher here compared to the hill-climbing attack in A.2.

It is considered that:

— the identification phase corresponds to find to right interface to the system (connecting a computer to the targeted signals, enabling the presentation of synthetic samples) and to get or write the optimization software for the synthetic sample generation. It is considered as easy in this example. The exploitation phase is just running the program to get access;

— elapsed time: 2 weeks for identification and 1 day for exploitation are realistic;

— expertise: Even if the attack method is known and published, setting up the right connections, exploiting specific signals and adapting an optimization software is considered to require an expert level for identification and a proficient level for exploitation;

— knowledge of the TOE: A restricted level of knowledge of the TOE is required;

— window of opportunity (access to the TOE): The TOE operates in a fully uncontrolled environment (rated as easy for exploitation phase) and it is considered as easy to buy system (rated as easy for identification phase);

— window of opportunity (access to biometric characteristics): It is rated as immediate as the attack does not require the availability of real data (synthetic templates);

— equipment: A specialized equipment is required (computer, connection to the system, synthetic samples generation, optimization software) in identification. For exploitation, the software being considered as available, the rating is standard.

With the above assumptions, the summary of the corresponding attack potential calculation is provided in Table A.3.

Table A.3 — Calculation of attack potential for the scenario in A.4

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				14	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
2	0	4	4	2	-	0	0	-	0	2	0	10	4

The rating for the attack is enhanced-basic.

If the attack can be performed successfully and no other successful attack with a lower rating is found, the resistance of the TOE is basic.

The system is compatible with the AVA_VAN.2 component.

A.5 Dictionary attack

EXAMPLE Assume the system is an iris, a two-irises or a face-based system operating in an uncontrolled environment (for example, protecting the access to a device or an equipment). There is an easy way to connect a computer just before the feature extractor. However, now it is not necessary to access to the comparison score. The attack consists in sending, by injection after the data capture subsystem, real biometric samples to the system until one is accepted.

It is considered that:

— the identification phase corresponds to find to right interface to the system (connecting a computer to the targeted signals, enabling the presentation of biometric images). This is considered as easy in this example. The exploitation phase is just inputting images to get access;

— elapsed time: 2 weeks for identification and exploitation are realistic;

— expertise: A proficient attacker should be able to locate the input of the feature extractor using some specialized equipment, and thus insert the sample images into the system;

— knowledge of the TOE: A restricted knowledge of the TOE is required;

— window of opportunity (access to the TOE): The TOE operates in an uncontrolled environment (rated as easy for exploitation phase) and it is considered as easy to buy system (rated as easy for identification phase);

— window of opportunity (access to biometric characteristics): The dictionary attack is conducted using real samples. Getting a large face database for the attack is very easy (rated as immediate) but getting a large enough iris database can be very difficult, and even more challenging for two irises. The rating in exploitation for access to biometric characteristics reflects both facts. On the other hand, note that although it can be difficult to get a large enough iris database, once obtained it can be used for attacking different system. The level easy reflects that fact. For two irises, for a similar reason, level moderate is selected;

— equipment: A specialized equipment is required (computer, connection to the system, data bases).

With the above assumptions, the summary of the corresponding attack potential calculation is provided in Table A.4 for the case of face, Table A.5 for the case of one iris, and Table A.6 for the two irises case.

Table A.4 — Calculation of attack potential for the scenario in A.5 (face)

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				16	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
2	4	2	0	2	-	0	0	-	0	2	4	8	8

Table A.5 — Calculation of attack potential for the scenario in A.5 (iris)

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				18	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
2	4	2	0	2	-	0	0	-	2	2	4	8	10

Table A.6 — Calculation of attack potential for the scenario in A.5 (two irises)

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				20	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
2	4	2	0	2	-	0	0	-	4	2	4	8	12

The rating for the attack is enhanced-basic (both face and iris) and moderate (for the two irises case).

If the attack can be performed successfully and no other successful attack with a lower rating is found, the resistance of the TOE is basic (both face and iris) and enhanced-basis (two irises).

The system is compatible with the AVA_VAN.2 (both face and iris) component and AVA_VAN.3 component (two irises).

NOTE The overall quotation of the attack would in principle depend on the FAR of the system, in particular for the elapsed time and for the equipment needed during exploitation. If it takes 1 second per attempt and if one needs 10 000 tries for FAR 10-3, it can take less than 2 weeks. Moreover, one can need a very large dataset for exploitation phase. For instance, for iris if operating point is FAR 10-6, the need of a dataset of more than 1 million of iris can be considered as a specialized equipment. For two irises, with a FAR below 10-8 for instance, the size of the dataset justifies increasing the equipment to bespoke, thus increasing the overall total for the attack to 26.

A.6 Wolf attack

EXAMPLE Assume the system is a biometric system for a given modality, operating in an uncontrolled environment (for example, protecting the access to a device or an equipment). Assume that it is straightforward to connect a computer just before the feature extractor (for instance because the TOE is purely software-based) in order to inject images as input of the feature extractor stage. Assume also that there is a flaw in the comparison algorithm such that there is a way to construct an image (not necessarily from natural biometric characteristics) for which the false acceptance rate is significantly higher than any randomly drawn biometric data. This corresponds to a system with a high rate of successful attack. The attack will consist in finding the particular (or one of) image(s) that leads to a high chance of being accepted.

It is considered that:

— the identification phase corresponds to find the weakness in the matching algorithm and thereafter to generate an image which exploits this flaw. The exploitation phase is just inputting the image to get access;

— elapsed time: More than one month for identification might be needed while exploitation is immediate;

— expertise: An expert attacker should be needed to find the flaw in the algorithm. A layman can input the image once generated;

— knowledge of the TOE: Sensitive knowledge of the TOE is required to learn the detail of the matching algorithm;

— window of opportunity (access to the TOE): The TOE operates in an uncontrolled environment (rated as easy for exploitation phase) and it is considered as easy to buy system (rated as easy for identification phase);

— window of opportunity (access to biometric characteristics): It is rated as immediate as no access is needed for the attack;

— equipment: A specific equipment may be required to generate an image, acceptable for the feature extractor.

With the above assumptions, the summary of the corresponding attack potential calculation is provided in Table A.7.

Table A.7 — Calculation of attack potential for the scenario in A.6

Elapsed time		Expertise		Knowledge of TOE		Window of opportunity				Equipment		Total	
						Access to TOE		Access to biometric characteristics				18	
Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp	Id	Exp
8	0	4	0	4	-	0	0	-	0	2	0	18	0

The rating for the attack is: enhanced-basic

If the attack can be performed successfully and no other successful attack with a lower rating is found, the resistance of the TOE is basic.

The system is compatible with the AVA_VAN.2 component.