ISO/IEC JTC 1/SC 27/WG 3 "Security evaluation, testing and specification"
Convenorship: **UNE**
Convenor: **Bañón Miguel Mr**

CEN/TC 224/WG 18 European requirements for biometric products Part 3 WD

Document type	Related content	Document date	Expected action
Project / Other		2024-03-27	**INFO**

Replaces: N 2589 CEN-TC 224-WG 18_N963_NP TS ERBP-3 WD5

CEN/TC 224/WG 18 "Biometrics"
WG Secretariat: **AFNOR**
Convenor: **Gacon Pierre M.**

TS ERBP-3 WD6

Document type	Related content	Document date	Expected action
Meeting / Document for discussion	Meeting: Berlin (Germany) 5 Jun 2024 Project: 00224279 - -	2024-03-22	**COMMENT/REPLY** by 2024-05-21

CEN/TC XXX

Date: 20XX -XX

(ERBP-3 WD6) **prEN XXXXX: XXXX**

Secretariat: XXX

Personal identification — European requirements for biometric products — Part 3: Functionality evaluation methodology

Einführendes Element — Haupt-Element — Ergänzendes Element

Élément introductif — Élément central — Élément complémentaire

ICS:

CCMC will prepare and attach the official title page.

Contents

Page

European foreword

This document (prEN XXXX:XXXX) has been prepared by Technical Committee CEN/TC XXX "Title", the secretariat of which is held by XXX.

This document is currently submitted to the CEN Enquiry.

This document has been prepared under a mandate given to CEN by the European Commission and the European Free Trade Association, and supports essential requirements of EU Directive(s).

For relationship with EU Directive(s), see informative Annex ZA, which is an integral part of this document.

[NOTE to the drafter: Add information about related documents or other parts in a series as necessary. A list of all parts in a series can be found on the CEN website.]

Introduction

The use of remote services has increased significantly. This was boosted during 2020-2021, when many service providers and Administrations migrated most of their processes to online handling. We can find nowadays many online services, such as opening of a bank account, claiming expenses, paying taxes, starting legal actions, etc.

For all these services there is the need of identifying the persons claiming for that service, and doing it in a comfortable, universal, reliable and auditable way. Even though some of those services, in some countries, were deployed using PKIs (Public Key Infrastructures), as recommended by eIDAS, this approach was far away from being used by a significant part of the population.

This situation led to creating identification services using videoconferencing tools, such as using any device camera to scan a document, and capture your face for biometric recognition. This is deployed in many countries and sectors, but using ad-hoc solutions, limiting interoperability and increasing costs and risks.

In this context, service providers and Administrations have to define their own requirements, select the products and deploy the solution. On the other hand, manufacturers had to implement different solutions to different customers, in order to fulfil each of those requirement sets. Both sides would benefit from standards and regulations, on which to rely for the product definition.

Everybody will benefit from having a common way of defining those requirements, and a detailed evaluation methodology. These two items can be used by conformity assessment bodies or by business owners, to create their own certification schemes for this kind of technology/products, by following the international ISO/IEC 17000 series of standards.

This project is addressing this need for the case of Biometric Products, analysing and merging all current works, and defining a detailed set of requirements, a biometric-mode-specific evaluation methodology, and the passing criteria for different application profiles. This work will be developed in accordance with GDPR principles.

This will be written as a multipart project with the following structure:

— Parts 1-3: Defining the generic principles and methodologies, not requiring a biometric mode specific approach. In particular these parts will be:

  — Part 1: General requirements and application profile definition

  — Part 2: Interoperability tests

  — Part 3: Functionality evaluation methodology

— Parts 4-n: Defining the particularities of each biometric mode (e.g., specific tests, specific requirements), and containing, each of the parts, a set of application profiles, that will establish the test and requirements applicable for a specific application and context. Those application profiles will be written as individual annexes, following the structure provided in Part 1. The numbering of these parts, has been done trying to keep conformance with the numbering used by ISO/IEC 19794 series of standards (ISO/IEC_JTC1/SC37_WG3). Therefore:

  — Part 4: Fingerprint biometrics

  — Part 5: Face biometrics

  — Etc.

NOTE FOR THE EDITOR: Figures shall fit the specifications from CEN. Apply that in the next cycle.

1 Scope

This TS series provide a generic framework for the establishment of requirements and their evaluation methodology for biometric products. The requirements will be established depending on the biometric mode considered, and they will be adapted to each scenario, through the definition of a variety of application profiles.

This series of standards are expected to provide the evaluation methodology, the individual tests, and the application profiles (with their particular requirements).

This document specifies:

— The different kind of evaluations to be performed

— The terms used during the description of the tests to be applied

— The parameters used, whose values will be defined by each application profile, for each of the individual tests

— Test data used, and considerations dealing with personal data protection

— How to perform technology evaluations

— Execution flow for functionality scenario evaluations

— Execution flow for attack resistance evaluations

NOTE Additional parts are provided covering the specifics of each biometric mode. For each of these modalities, application-independent tests are defined, as well as a set of application profiles, that detail the applicable tests, the evaluation parameters, and the passing criteria.

The Technical Specifications within this series can be taken by any certification body and/or sector, to define and evaluate the requirements for their biometric products within their selected applications. This may be used in coordination with other current National initiatives. For governmental applications, the relevant Government will decide if this evaluation is applicable or not.

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.

CEN TS (WI=00224273) *Biometric data injection attack detection*

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

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

ISO/IEC 30107-1, *Information technology — Biometric presentation attack detection — Part 1: Framework*

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

ISO/IEC 30108 (all parts), *Biometrics – Identity attributes verification services*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in the first part of this series, ISO/IEC 19795 series, ISO/IEC 30107 series, ISO/IEC 2382-37 and the following apply.

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

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

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

NOTE Certain terms, being common-use words, are used in capitals throughout the text to make it clear for the reader that they are evaluation parameters, not regular terms.

3.1 General terms

3.1.1
Biometric functionality subsystem
set of software modules that perform the biometric functions within the biometric product.

Note 1 to entry: Examples of biometric functions are quality checking, pre-processing, comparison, presentation attack detection methods.

3.2 Evaluation elements and parameters

3.2.1
ARTIFACT
Artificial object or representation, that present a copy of the biometric characteristics of a SUBJECT

3.2.2
ATTACKER
Person that attacks the system. It can be an imposter of use an ARTIFACT for attempting a presentation attack

3.2.3
ATTEMPT
Each of the individual interactions between the SUBJECT and the TOE within a TRIAL

3.2.4
ERROR
Situation in which the TOE is not able to operate correctly, and therefore, is not able to accomplish a result of the biometric comparison

Example 1 to entry: The TOE is not able to acquire a biometric sample from a bona-fide SUBJECT due to low quality samples captured.

Note 1 to entry: In the case of a PAD TEST, an ERROR (once the maximum number of allowed ATTEMPTS has been reached) can be considered as a NON-MATCH, as the ARTIFACT was not able to be properly captured.

3.2.5

FAIL

For those TESTs within Phase 2 and Phase 3, it is the final result for such TEST, which tells that the TOE behaviour is not appropriate. On the other hand, during Phase 4, a result of FAIL, tells that the attack has not been successful and, therefore, the TOE behaviour is the correct one

3.2.6

MATCH

Positive result of a biometric comparison during a TRIAL

Example 1 to entry: A bona-fide SUBJECT acceptance in a functional TEST.

Note 1 to entry: In the case of a PAD TEST, a MATCH is the non-desired result, as it will show that the ARTIFACT used was able to achieve a successful comparison.

3.2.7

NON-MATCH

Negative result of a biometric comparison during a TRIAL

Example 1 to entry: A bona-fide SUBJECT rejection in a functional TEST.

Note 1 to entry: In the case of a PAD TEST, a NON-MATCH is the desirable result, as it will show that the comparison with the ARTIFACT used was not successful.

3.2.8

OPERATOR

Human being that, based on the TOE acquired data and result, take the decision on whether the transaction is valid or not

3.2.9

PASS

For those TESTs within Phase 2 and Phase 3, it is the final result for such TEST which tells that the TOE is presenting an appropriate behaviour. On the other hand, during Phase 4, a result of PASS, tells that the attack has been successful and, therefore, the TOE is vulnerable

3.2.10

SERVER

Computer-based equipment in which the TOE stores the acquired data during the biometric recognition process. Such data can be analysed later by an OPERATOR

3.2.11

SETTING

Execution context for a TRIAL within a TEST. The SETTING can be the description of equipment to use, the way the SUBJECT has to interact with the TOE, ambient conditions, ARTIFACTs to be used, etc. For each TEST, one or several SETTINGs have to be specified

3.2.12

SUBJECT

Individual whose biometric data is intended to be enrolled or compared as part of the evaluation. Traditionally, a SUBJECT will be a USER, but in certain evaluations the SUBJECT is a combination of a USER and some additional property or element

Example 1 to entry: In the case of a videoconference system, where the TOE is being used with a USER a potentially a variety of documents, the SUBJECT will be the combination of USER plus document.

Example 2 to entry: In the case Phase 4 test, a SUBJECT is the combination of USER, ARTIFACT, and any other relevant property.

3.2.13

TEST

Action to evaluate the behaviour of the TOE for certain features. One TEST is composed of several TRIALS, which involve several SUBJECTS and, probably, several SETTINGs

3.2.14

TEST_ERROR

Situation in which, within a TEST, the TRIALs corresponding to a certain SUBJECT get over the limit MAX_SUBJECT_ERRORS

3.2.15

TRIAL

Each of the interactions between the SUBJECT and the TOE, during the TEST. Depending on the TOE, each TRIAL may allow several ATTEMPTS

Example 1 to entry: The TOE may ask the SUBJECT to repeat the biometric presentation due to acquisition errors. In such a case, the new presentation will be considered as a new ATTEMPT within the same TRIAL.

3.2.16

USER

Human being that takes part in a TRIAL. Depending on the TEST, the USER could be a bona-fide SUBJECT or an ATTACKER, or it can behave in one TRIAL as a bonafide SUBJECT, and in another TRIAL as an ATTACKER

4 Symbols and abbreviations

4.1 General symbols and abbreviation

CSA Cybersecurity Act (Comission, n.d.)

eIDAS electronic Identification, Authentication and Trust Services

ETR Evaluation Technical Report

EU European Union / European

GDPR General Data Protection Regulation

ID Identity

LoA	Level of Assurance
PAD	Presentation Attack Detection (as described in ISO/IEC 30107-1)
TL	Testing Laboratory
TOE	Target of Evaluation

4.2 Symbols related to the evaluation workflow

— **MAX_ATTEMPTS:** Maximum number of ATTEMPTS allowed for a TRIAL, before resulting in an ERROR for that TRIAL.

— **MAX_SETTING_MATCHES:** Maximum number of TRIALS, among all required for a SETTING during a TEST, that provide a MATCH result. When this number is reached, the TEST is considered as a PASS for that SETTING. This is only applicable to Phase 4.

NOTE In a Phase 4 TEST, a PASS result means that the TOE is vulnerable for that attack.

— **MAX_SETTING_NON_MATCHES:** Maximum number of TRIALS, among all required for a SETTING during a TEST, that provide a NON-MATCH result. When this number is reached, the TEST is considered as FAIL for that SETTING. This is only applicable to Phases 2 and 3.

— **MAX_SUBJECT_ERRORS:** Maximum number of ERRORs allowed for the sum of all TRIALS for a single SUBJECT, within a particular SETTING and TEST. This is only applicable to Phases 2 and 3.

— **MAX_SUBJECTS_FAIL:** Maximum number of SUBJECTs, for which TRIALS within a SETTING and TEST have reached the limit of MAX_SUBJECT_NON_MATCHES. This is only applicable to Phases 2 and 3.

— **MAX_SUBJECT_MATCHES:** Maximum number of TRIALS with a MATCH result, allowed for a single SUBJECT within one SETTING. This is only applicable to Phase 4.

NOTE In a Phase 4 TEST, a PASS result means that the TOE is vulnerable for that attack.

— **MAX_SUBJECT_NON_MATCHES:** Maximum number of TRIALS with a NON-MATCH result, allowed for a single SUBJECT within one SETTING. This is only applicable to Phases 2 and 3.

— **MAX_SUBJECTS_PASS:** Maximum number of SUBJECTs, for which TRIALS within a SETTING and TEST have reached the limit of MAX_SUBJECT_MATCHES. This is only applicable to Phase 4.

— **MAX_TEST_ERRORS:** Maximum number of SUBJECTS, within a TEST, for which its TRIALS have reached the limit given by MAX_SUBJECT_ERRORS. This is only applicable to Phases 2 and 3.

— **MAX_TEST_MATCHES:** Maximum number of TRIALS, among all included in a TEST, with a MATCH result. If such number is reached, the TEST is considered as a PASS. This is only applicable to Phase 4.

NOTE In a Phase 4 TEST, a PASS result means that the TOE is vulnerable for that attack.

— **MAX_TEST_NON_MATCHES:** Maximum number of TRIALS, among all included in a TEST, with a NON-MATCH result. If such number is reached, the TEST is considered as FAIL. This is only applicable to Phases 2 and 3.

— **MIN_SETTINGS:** Minimum number of SETTINGS defined.

— **MIN_SUBJECTS:** Minimum number of SUBJECTS defined.

— **MIN_TRIALS:** Minimum number of TRIALS defined.

5 General concepts

5.1 General

As explained in part 1 of this series, the evaluation of a biometric product is done through 4 phases, where the first one, detailed in part 2, is focused on the interoperability aspects relevant to the TOE and the application profile. But phases 2 to 4 are focused on evaluating the biometric functionality of the TOE, regarding performance, suitability to the application profile and robustness against presentation attacks.

This document defines the basis for all the functional evaluation, i.e. the tasks to execute phases 2 to 4. This functional evaluation is based on the specifications provided by ISO/IEC 19795 series and ISO/IEC 30107 series.

All parts in this series beyond this 3rd part, specify the biometric mode-specific tests to be executed, as well as a set of application profiles. Each of those application profiles will determine the main characteristics of the TOE for which the application profile is applicable, as well as which are the applicable tests, and the acceptance criteria for each of the tests, as well as for the overall functional evaluation.

In order to better understand the general testing methodology, this clause will revisit the evaluation phases introduced in part 1, as well as the relationship with both families of ISO/IEC standards mentioned above.

After this clause, clause 6 will present important fact dealing with the test data. Clause 7 will define the methodology for phases 2 and 3, and clause 8 will specify the methodology to be followed in phase 4 tests.

Last, but not list, clause 9 will specify the additional methodology that shall be applied on top of the one provided in clauses 7 and 8, for those cases where the biometric functionality subsystem of the TOE has been developed using Machine Learning tools.

5.2 Functional evaluation phases

Within this conformity assessment methodology, the evaluation of the TOE, shall be done following the phases defined in Part 1 of this series of standards. This document is focussed on the definition of Phases 2 to 4, which are expected to be executed in a sequential manner:

— Phase 2: TOE performance tests

  — The main target of these TESTs is to verify the TOE behaviour regarding what it has been declared by the product supplier. This is to be checked using the relevant SETTINGs for the application profile selected.

— Phase 3: Bona-fide robustness tests

  — The main target of these TESTs is to learn about the TOE, as to be able to locate the operating boundaries in using the TOE with bona-fide SUBJECTs.

  — This knowledge may help evaluators to discover strategies to attack the TOE during Phase 4 tests.

  - Results obtained will be checked with the TOE documentation, as to check is the FAILed tests are clearly excluded from the TOE usage.

- Phase 4: Presentation attack detection tests

  - The main target of these tests is to determine if the TOE is vulnerable to presentation attacks, either Type 1 or Type 2 attacks (as defined in ISO/IEC 30107-1 and CEN TS Digital Injection).

  - According to the application profile, the evaluated attacks may be impostor attacks, concealer attacks or both.

  - The EU Cybersecurity Act (EUCSA, Regulation 2019/881 (Union, 2019)) defines 3 levels of assurance (LoA), named as Basic, Substantial and High.

  - Under a LoA "High" (as defined by the EU Cybersecurity Act – EUCSA), any Phase 4 ATTEMPT resulting in a PASS, will declare a FAIL for the biometric product to achieve such LoA. This will be determined by analysing that the attack is not exceeding the maximum attack potential for the TOE evaluation.

5.3 Compliance with ISO/IEC 19795 series

Phases 2 and 3 evaluate the performance and suitability of the TOE for the application profile defined. To achieve that evaluation, ISO/IEC 19795 present the basis. ISO/IEC 19795 is titled "Information technology — Biometric performance testing and reporting" and is a multipart standard, where the most relevant parts are:

- Part 1: Principles and framework

- Part 2: Testing methodologies for technology and scenario evaluation

- Part 3: Modality-specific testing

- Part 9: Testing on mobile devices

Part 9 shall be considered when the TOE is a mobile device. Relevant clauses from Part 3 shall be used as an input to each of the biometric-mode-specific parts of this standard series. Parts 1 and 2 define all the evaluation principles and the basic testing methodology.

Within these principles, three kinds of evaluations are defined:

- Technology evaluations: where testing is carried out on a standardized corpus, ideally collected by a "universal" sensor. In other words, this kind of evaluation is thought to be applied directly to the biometric algorithm, and using a previously collected database.

- Scenario evaluations: where testing is carried out on a complete system in an environment that models a real-world target application of interest. So, the evaluation is performed using real subjects (i.e., not a database), where the context in which the TOE is expected to be used, is simulated at the TL.

- Operational evaluations: that are those evaluations when the TOE is deployed in the real application, and the evaluation is being performed while the system is under its expected operation.

Within this standardization series, operational evaluations are not considered. Most of the tests defined are scenario-based tests, but some others will use databases, approaching the concept of a technology evaluation.

5.4 Compliance with ISO/IEC 30107 series

Phase 4 is focused on evaluating the robustness of the TOE under those relevant attacks. Most of those attacks are Presentation Attacks, as defined in ISO/IEC 30107-1. For the evaluation of the capability of Presentation Attack Detection (PAD), ISO/IEC 30107-3 define the general methodology in a biometric mode agnostic manner, specifying the basis for a more detailed and applicable methodology.

Therefore, PAD tests in Phase 4 shall use ISO/IEC 30107-3 as the initial specification of the evaluation. Also, when reporting the results, ISO/IEC 30107-3 shall be followed. ISO/IEC 30107-3 define two main philosophies for carrying out PAD evaluation. When the relevant application profile requires a LoA “High” or “Substantial”, the Common Criteria approach shall be used, which is detailed in the clause titled “Evaluation using Common Criteria framework”.

5.5 Terms and parameters used during the evaluation

Most of biometric TESTs follow a very similar execution sequence, which is described in clauses 7 and 8. Such clauses are written in a generic way, so as to allow an easier description of each of the TESTs. Other parts of this Technical Specification will define each of the specific TESTs, based on that sequence.

For a better understanding of this methodology, the following terms are needed (defined in clause 3):

— ARTIFACT

— ATTACKER

— ATTEMPT

— ERROR

— FAIL

— MATCH

— NON-MATCH

— OPERATOR

— PASS

— SERVER

— SETTING

— SUBJECT

— TEST

— TEST_ERROR

— TRIAL

— USER

It is also important to consider the following parameters that will be used throughout this evaluation methodology (defined in clause 4):

— MAX_ATTEMPTS

— MAX_SETTING_MATCHES

— MAX_SETTING_NON_MATCHES

— MAX_SUBJECT_ERRORS

— MAX_SUBJECTS_FAIL

— MAX_SUBJECT_MATCHES

— MAX_SUBJECT_NON_MATCHES

— MAX_SUBJECTS_PASS

— MAX_TEST_ERRORS

— MAX_TEST_MATCHES

— MAX_TEST_NON_MATCHES

— MIN_SETTINGS

— MIN_SUBJECTS

— MIN_TRIALS

MIN_TRIALS, MIN_SETTINGS and MIN_SUBJECTS define the minimum number specified for each TEST. These are the numbers to be used by the TL. If during an evaluation the TL detects too many ERRORs during the TRIALS, the TL may increase those numbers, until it can obtain a number of conclusive (i.e., NON-ERROR) results, equal to:

$$Minimum\,conclusive\,results = MIN_{TRIALS} * MIN_{SETTINGS} * MIN_{SUBJECTS} \quad (1)$$

This deviation shall be fully justified and included in the ETR.

6 Test data

6.1 General considerations

Data is needed for performing biometric evaluations.

In the case of technology evaluations, when the biometric capture subsystem can be detached from the TOE, previously recorded databases can be used to speed up the evaluation, increasing also the significance of the results obtained.

In those cases where the biometric capture subsystem cannot be detached from the rest of the TOE, test data can only be obtained by calling test crews.

GDPR has always to be preserved

6.2 Stored databases

6.2.1 Recorded databases

Most of the tests to be defined under this evaluation methodology are going to be scenario-based tests, which means using real users as input to the TOE. But there are some tests that can be considered as technology evaluations and, therefore, use databases.

For those kind of tests, ISO/IEC 19795 parts 1 and 2 shall be followed. As it is stated in those two standards, the databases shall be representative of the target population where the TOE is going to be applied, and be varied enough as to be able to cover most of the diversity of such population.

Databases can be previously recorded and used in several evaluation of TOEs, as long as GDPR regulation is followed, and the representativeness of the database is guaranteed.

NOTE As a minimum requirement for GDPR, the records within a database shall be anonymized whenever possible.

The above-mentioned technology tests are typical from either interoperability testing (see part 2), or for some of the tests in Phase 2 and Phase 3 (see clause 7).

6.2.2 Use of synthetic databases

Due to the difficulty of creating large databases, plus the challenges of applying GDPR to such database, it may be considered the use of synthetic databases, which will not be impacted by GDPR, and may remove the challenge of achieving a large number of biometric samples.

But in order to use such synthetic database, it has to be proven that:

— The database is representative of the target population indicated by the application profile. In order to reach this objective, the distribution of the database shall be representative in terms of gender, age, ethnicity, and/or any other relevant parameter important for the target population.

— The database samples shall be realistic enough, so that the behaviour of state-of-the-art algorithms may be considered equivalent with the performance achieved using real sample databases.

In order to achieve this second requirement, synthetic biometric samples shall be noisy enough, as to allow the algorithms to achieve, not only an equivalent value of FNMR@FMR, but also an equivalent FTA rate.

EXAMPLE In the case of fingerprint mode, there is a well-known application called SFinGe (Laboratory, n.d.), which can create completely clean fingerprints. In version 2 they started to add image distortion methods, to keep into account skin plasticity. In version 2.5 they added the generation of realistic backgrounds and different fingerprint sizes. And from version 3.0 till now, they have added improved noising algorithms and parameters. In the current version 5, they have even added a parameter to control the probability of generating very-low quality fingerprints.

In order to determine that the synthetic database can be used for the evaluations defined by this document, 5 state-of-the-art biometric algorithms shall be used.

The testing set shall be dimensioned and composed in such a a manner that it can provide a statistical relevant representation to the performance on the societal clusters composed on Age, Gender, Ethnicity, labour status, and/or any further criteria relevant to the application profile.

Each of those algorithms shall be executed against a real dataset and the synthetic data set. The real dataset shall be also equivalent to the synthetic database, i.e., representative of the target population.

NOTE Depending on the relevant biometric mode, there could be repositories (or listings) of state-of-the-art algorithms. Each of the biometric-mode-specific part (i.e., Parts 4-n) of this series of standards, may define the repository of algorithms to be used for this task.

For the application profile requested FMR, both FNMR and FTA rate shall provide numbers within the same order of magnitude between the execution with the real dataset and the synthetic dataset.

If the results show that for at least 4 out of the 5 algorithms, both error rates are within the same order of magnitude, the synthetic database shall be considered valid for being used in the evaluation defined in this series of standards.

6.3 Test crews in scenario evaluations

When a scenario-based test is required, then databases are not used, but real users. The use of human beings as test crew members at the moment of the evaluation, drives important challenges to the TL, specially if that same test subject is expected to show into the evaluation several times.

The application profile may consider to limit the test crew size, as to allow a higher viability of the evaluation, in particular when the evaluation shall face some time and/or cost limitations.

Some tests may impose particular features of the test crew members, such as diversity in the biometric characteristics, similarity among them, possibility of acting in different manners when interacting with the TOE, etc. For sure, they will have to be careful in following all indications given by the TL members.

In Phase 2, test crew members shall not be involved in the evaluation of the TOE as to keep a behaviour not biased by an excessive knowledge of the TOE. In Phases 3 and 4, this requirement is also recommended.

At all moment, GDPR shall be respected, to protect the privacy of each of the test crew members. When a sample from a test crew member is needed to explain the results in the ETR, that sample shall be anonymized as much as possible by, for example, segmenting all non-significant information within the sample.

7 Evaluation process for Phase 2 and 3

7.1 Overall view of the scenario evaluation

At Phases 2 and 3, several scenario evaluations are executed. For this methodology, each of these evaluations is called a TEST. Each TEST will consider a number of SETTINGS and a set of SUBJECTS (i.e., a test crew).

For each combination of SETTINGS and SUBJECTS, a number of TRIALS will be performed, being possible that each TRIAL allows a maximum number of ATTEMPTS.

The following figure represents the hierarchical relationship among these elements.

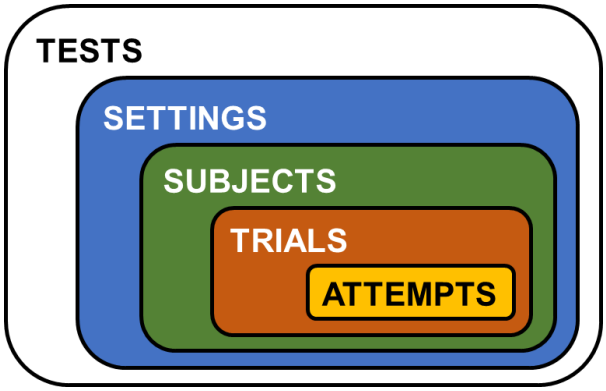

Figure 1 – Hierarchical relationship among evaluation elements

This will be the hierarchical relationship that will be used during the whole description of this evaluation methodology.

But, depending on the evaluation, it could be interesting to exchange the order among TESTS, SETTINGS and SUBJECTS. For example, the TL might consider more practical to execute all TESTs relevant to the same SETTING to all SUBJECTS, before changing the SETTING. Or it could be more practical to execute all TEST with all SETTINGS for each of the SUBJECTS. This decision is up to the TL. If the relationship given in Figure 1 is modified in any manner, this shall be justified and detailed in the ETR. The following figure show some alternatives.

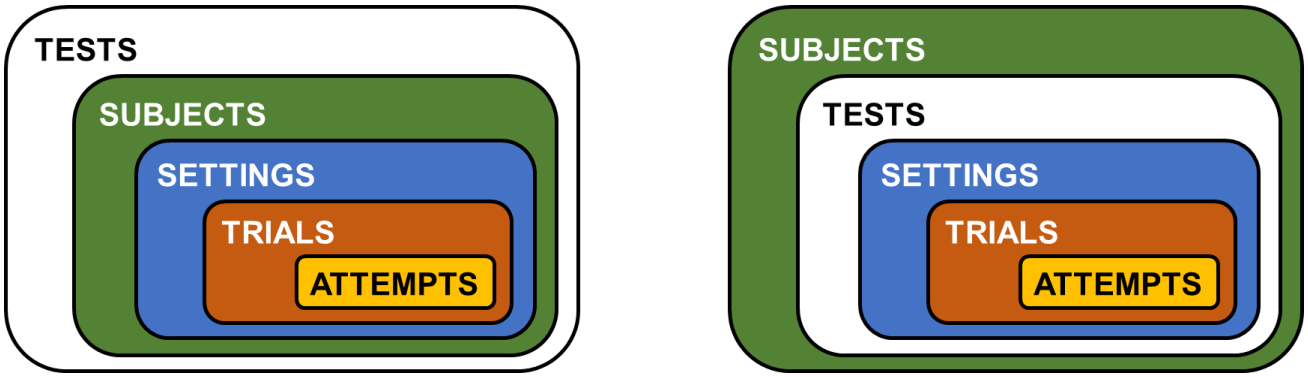

Figure 2 – Alternatives to the relationship among evaluation elements

7.2 TEST-level process

Each TEST is composed by the execution of a series of SETTINGS, up to reaching the limit given by MIN_SETTINGS. The TL may increment this number if the number of conclusive results is below the one demanded by equation (1).

Once the execution of all SETTINGS, for all SUBJECTS and TRIALS is finished, the final results are analysed to determine if the TEST is a PASS or FAIL, according to the criteria provided by the relevant application profile.

The flowchart for the TEST-level process is given in the following figure, which includes how to handle the SETTINGS.

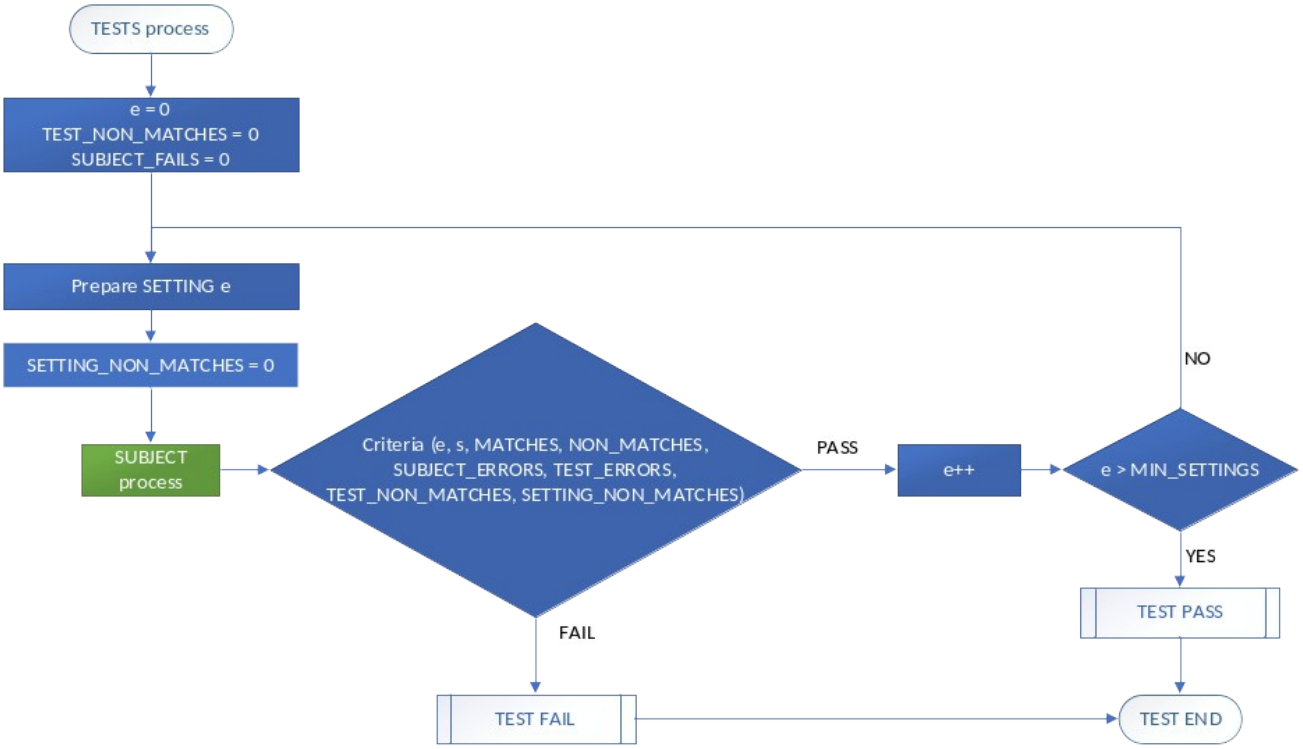

Figure 3 – Flowchart for the TEST-level process (Phases 2 and 3)

7.3 SUBJECT-level process

During the execution of each SETTING, several SUBJECTS take part until, at least, a number of MIN_SUBJECTS is reached. The TL may increment this number if the number of conclusive results is below the one demanded by equation (1).

Once the execution of all TRIALS for each SUBJECT is finished, the number of SUBJECT_ERRORS, MATCHES and NON_MATCHES obtained are analysed.

When all SUBJECTS have been evaluated, the final results for all SUBJECTS are analysed, to determine if the SETTING is a PASS or a FAIL. This will be done if a FAIL has not been declared before finishing with all SUBJECTS.

The flowchart for the SUBJECT-level process is given in the following figure, calling the TRIALS-level process:

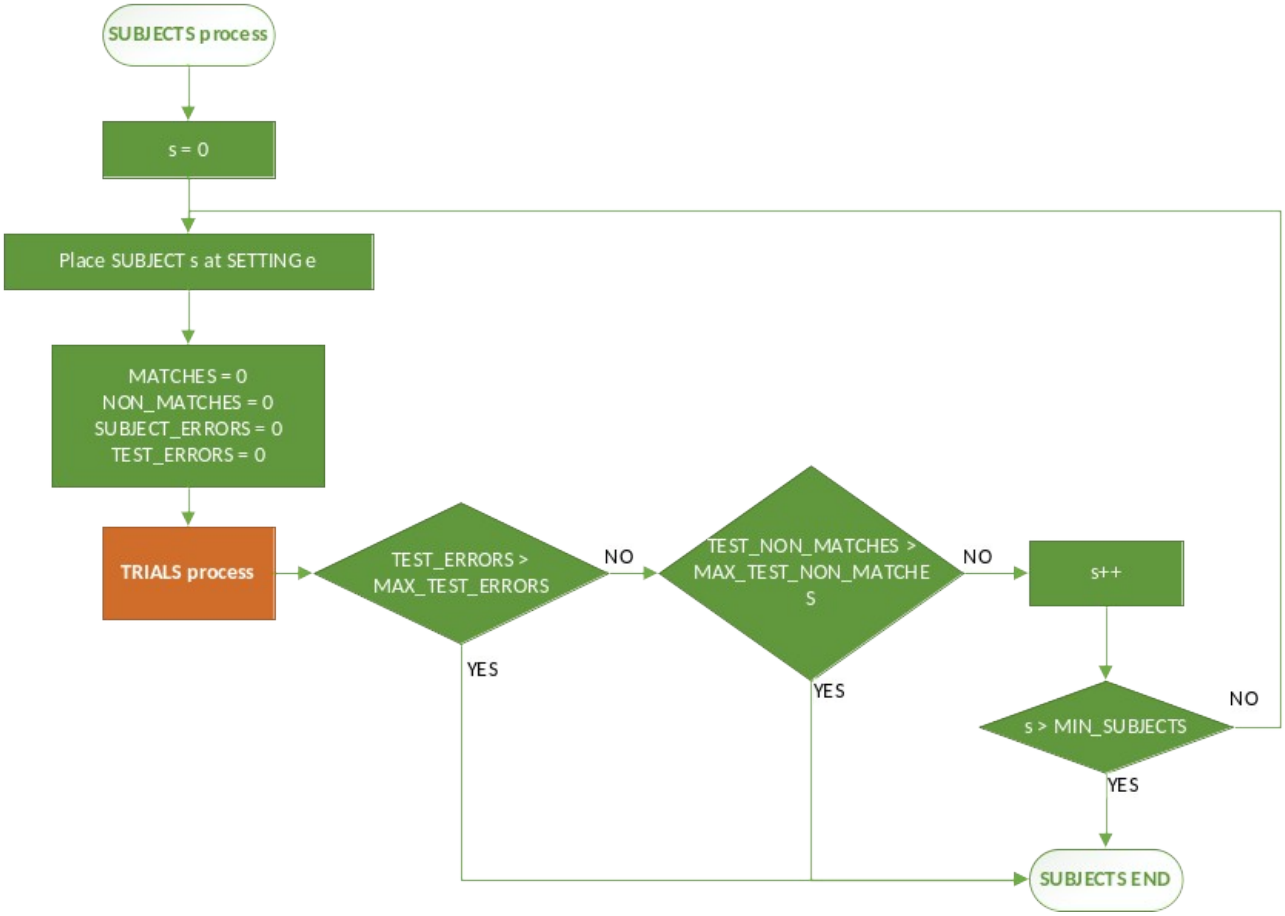

Figure 4 – Flowchart for the SUBJECT-level process (Phases 2 and 3)

7.4 TRIAL-level process

For each SUBJECT, a set of TRIALS are executed, until, at least, a number of MIN_TRIALS is reached. The TL may increment this number if the number of conclusive results is below the one demanded by equation (1). During the execution of each of the TRIALS, several ATTEMPTS may be allowed (up to the maximum limit given by MAX_ATTEMTPS), until either a MATCH, NON-MATCH or ERROR is obtained.

If when executing an ATTEMPT the TOE does not offer a result, but fails in its execution, a new ATTEMPT will be started. This will be done until the limit of MAX_ATTEMPTS is found. If such limit is reached, the TRIAL will result in an ERROR, and the TEST_ERRORS counter will be incremented.

If the number of TRIALS resulting in ERROR reaches the limit given by MAX_SUBJECT_ERRORS, TRIALS will be finished for that SUBJECT.

If during the whole TEST, the number of TEST_ERRORS reaches the limit given by MAX_TEST_ERRORS, this will be reported at the ETR. Then, the TL will add a new SUBJECT, and all TRIALS are executed for that new SUBJECT, decreasing TEST_ERRORS in one unit. If this situation is repeated, the TEST will be finished with a FAIL result.

In each ATTEMPT, if the result is a MATCH, the counters MATCHES and TEST_MATCHES are be incremented, and a new TRIAL is started. In case the result is a NON-MATCH, the counters NON_MATCHES,TEST_NON_MATCHES and SETTING_NON_MATCHES are incremented in one unit, and a new TRIAL is started.

If the number of NON_MATCHES is higher than MAX_SUBJECT_NON_MATCHES, the TEST will be finished for that SUBJECT, indicating a TEST FAIL for that SUBJECT. In such a case, the TEST continues with the following SUBJECT.

If the number of TEST_NON_MATCHES reaches the limit given by MAX_TEST_NON_MATCHES, the TEST will be finished, applying the defined criteria for that situation in the relevant application profile.

It is very important to consider that, in each ATTEMPT, the SUBJECT has to interact with the TOE, in the way it is indicated in the operational guide given by the product supplier. In other words, between ATTEMPTS, the SUBJECT shall withdraw from the interaction with the TOE in a significant manner.

EXAMPLE In the case of a videoconference system, the SUBJECT shall move temporally away from the focus line of the TOE, before returning for the new ATTEMPT.

The flowchart for the TRIAL-level process is given in the following figure, including the ATTEMPTS:

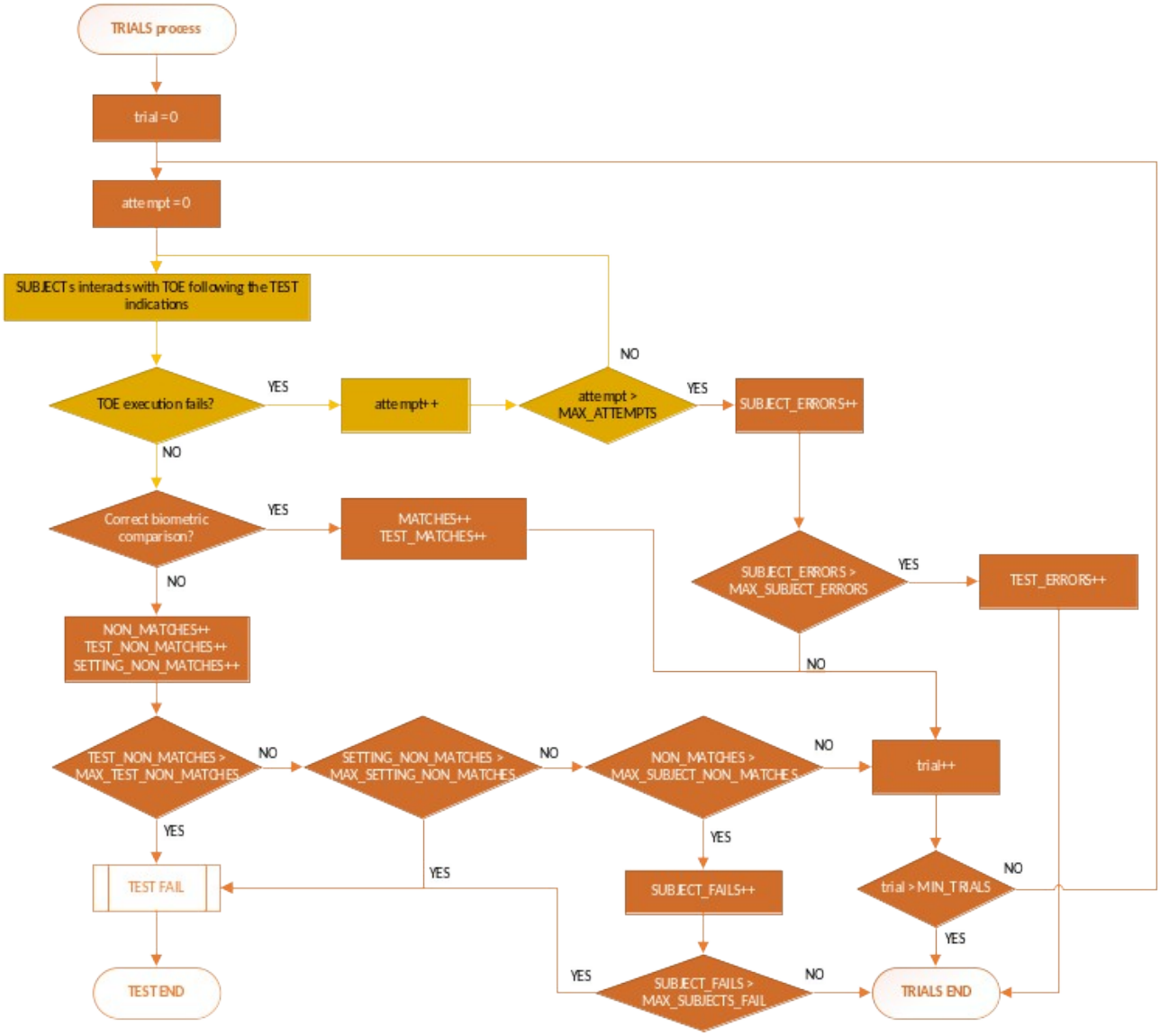

Figure 5 – Flowchart for the TRIAL-level process (Phases 2 and 3)

7.5 Families of tests in Phase 2

Many different tests can be defined for Phase 2, and this will be detailed in parts 4-n, for each biometric mode.

As a general rule, tests in Phase 2 can be grouped in the following set of families:

— Technology evaluation of the TOE, as to analyse the base-line performance under bona-fide and regular conditions.

— Operation in the recognition considering regular variations of the conditions of the SUBJECT (e.g., facial expression, finger humidity, etc.)

— Operation in the recognition considering regular variations of the scenario, i.e., the SETTING (e.g., environment illumination, background scenery, etc.)

7.6 Families of tests in Phase 3

Many different tests can be defined for Phase 3, and this will be detailed in parts 4, or beyond, for each biometric mode.

As a general rule, tests in Phase 3 can be grouped in the following set of families:

— Limits in the recognition considering sensible variations of the conditions of the SUBJECT (e.g., facial expression, finger humidity, etc.)

— Limits in the recognition considering sensible variations of the scenario, i.e., the SETTING (e.g., environment illumination, background scenery, etc.)

8 Evaluation process for Phase 4

8.1 Overall view of the scenario evaluation

The description of the scenario evaluation for Phase2 1 and 2 (i.e., clause 7.1) is also applicable to Phase 4. But the most important difference, is that in Phase 4, a TEST resulting in a PASS, means that the TOE is vulnerable for that TEST, and therefore, the desired result for Phase 4 TESTS is FAIL.

8.2 TEST-level process

For the TEST-level process, the following figure represent its flow chart, which calls the SUBJECT-level process:

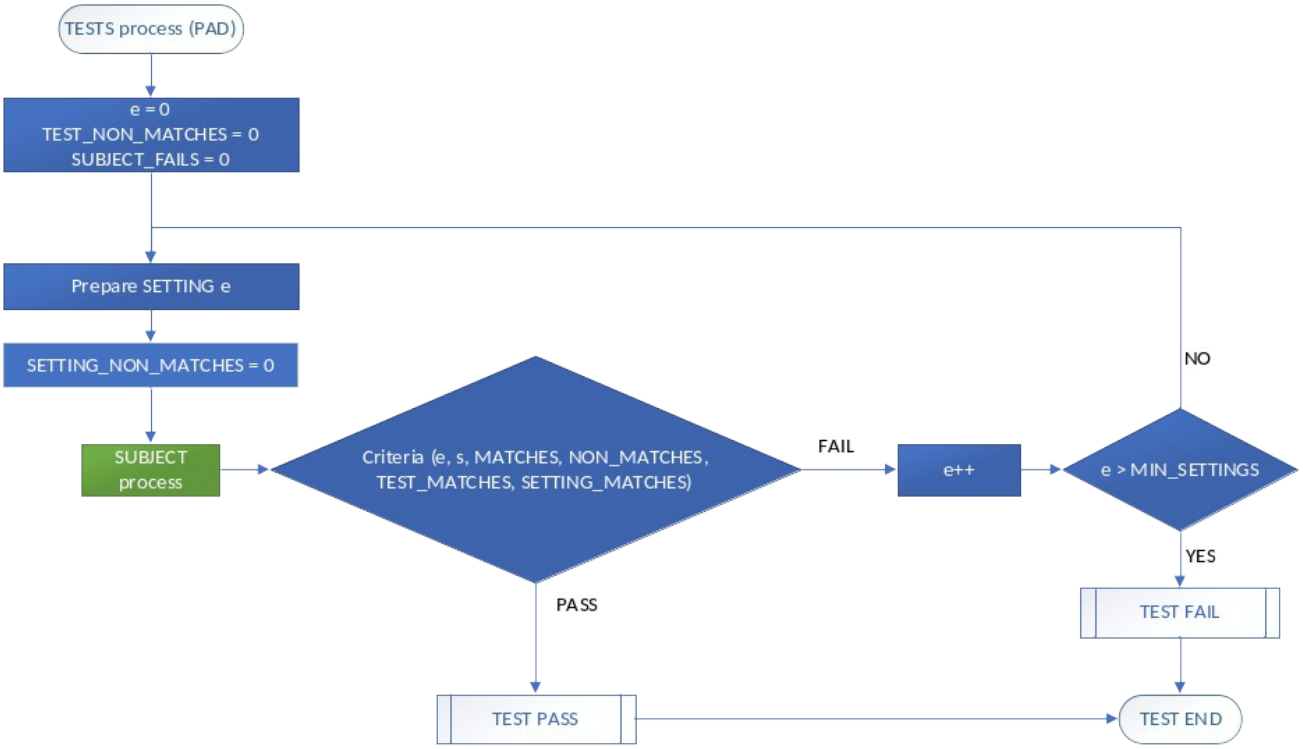

Figure 6 – Flowchart for the TEST-level process (Phase 4)

8.3 SUBJECT-level process

In Phase 4, during the execution of the SETTING process, several BONA-FIDE USERS will take part, until the minimum of MIN_SUBJECTS is reached. The TL may increment this number if the number of conclusive results is below the one demanded by equation (1).

Once the execution of all TRIALS for each SUBJECT is finished, the number of MATCHES and NON_MATCHES obtained are analysed for that SUBJECT. Once all SUBJECTS have gone through the TEST, the results will be analysed as to decide if the TEST results in a PASS or a FAIL

The following figure shows the flowchart for the SUBJECT-level process, which calls the TRIAL-level process:

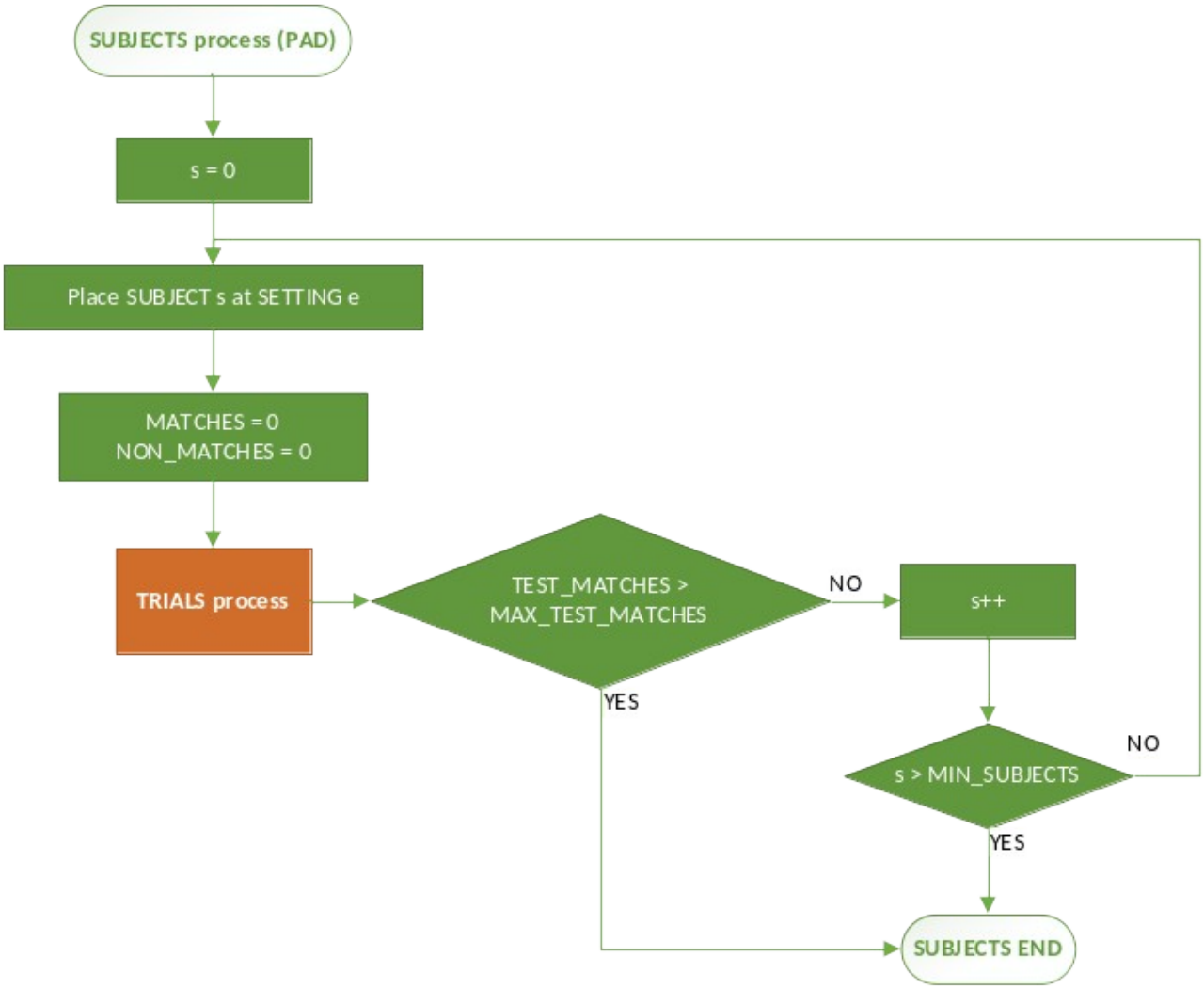

Figure 7 – Flowchart for the SUBJECT-level process (Phase 4)

8.4 TRIAL-level process

In Phase 4, a series of TRIALS are executed with each SUBJECT, until the minimum of MIN_TRIALS is reached. The TL may increment this number if the number of conclusive results is below the one demanded by equation (1). During the execution of each TRIAL, several ATTEMPTS may be allowed (until the limit of MAX_ATTEMPTS is reached). For each TRIAL, a result of MATCH, NON_MATCH or ERROR, shall be obtained.

When executing an ATTEMPT, if the TOE does not provide neither a MATCH, nor a NON_MATCH, a new ATTEMPT will be executed. This will be done until the limit of MAX_ATTEMPTS is reached. If that limit is reached, the TRIAL results in an ERROR. In Phase 4, an ERROR is a desirable result, as it tells that the TRIAL was not successful and, therefore, a PASS has not achieved.

If the TOE results in a MATCH, the counters MATCHES, TEST_MATCHES and SETTING_MATCHES are incremented. In case the result is a NON_MATCH, the counters to increment are NON_MATCHES and TEST_NON_MATCHES.

If the number of MATCHES reaches the limit given by MAX_SUBJECT_MATCHES, the TEST for that SUBJECT is finished, and a PASS will be assigned to such TEST for that SUBJECT. The next SUBJECT starts the TRIAL

If during the TEST execution, the number of TEST_MATCHES reaches the limit of MAX_TEST_MATCHES, the TEST will be finished, indicating a PASS for that TEST. Also, if during the TEST, the number of SETTING_MATCHES reaches the limit given by MAX_SETTING_MATCHES, the TEST is finished with a PASS as a result.

It is very important to consider that, in each ATTEMPT, the SUBJECT has to interact with the TOE, in the way it is indicated in the operational guide given by the product supplier. In other words, between ATTEMPTS, the SUBJECT shall withdraw from the interaction with the TOE in a significant manner.

EXAMPLE In the case of a videoconference system, the SUBJECT shall move temporally away from the focus line of the TOE, before returning for the new ATTEMPT.

The flowchart for the TRIAL-level process is given in the following figure, including the ATTEMPTS:

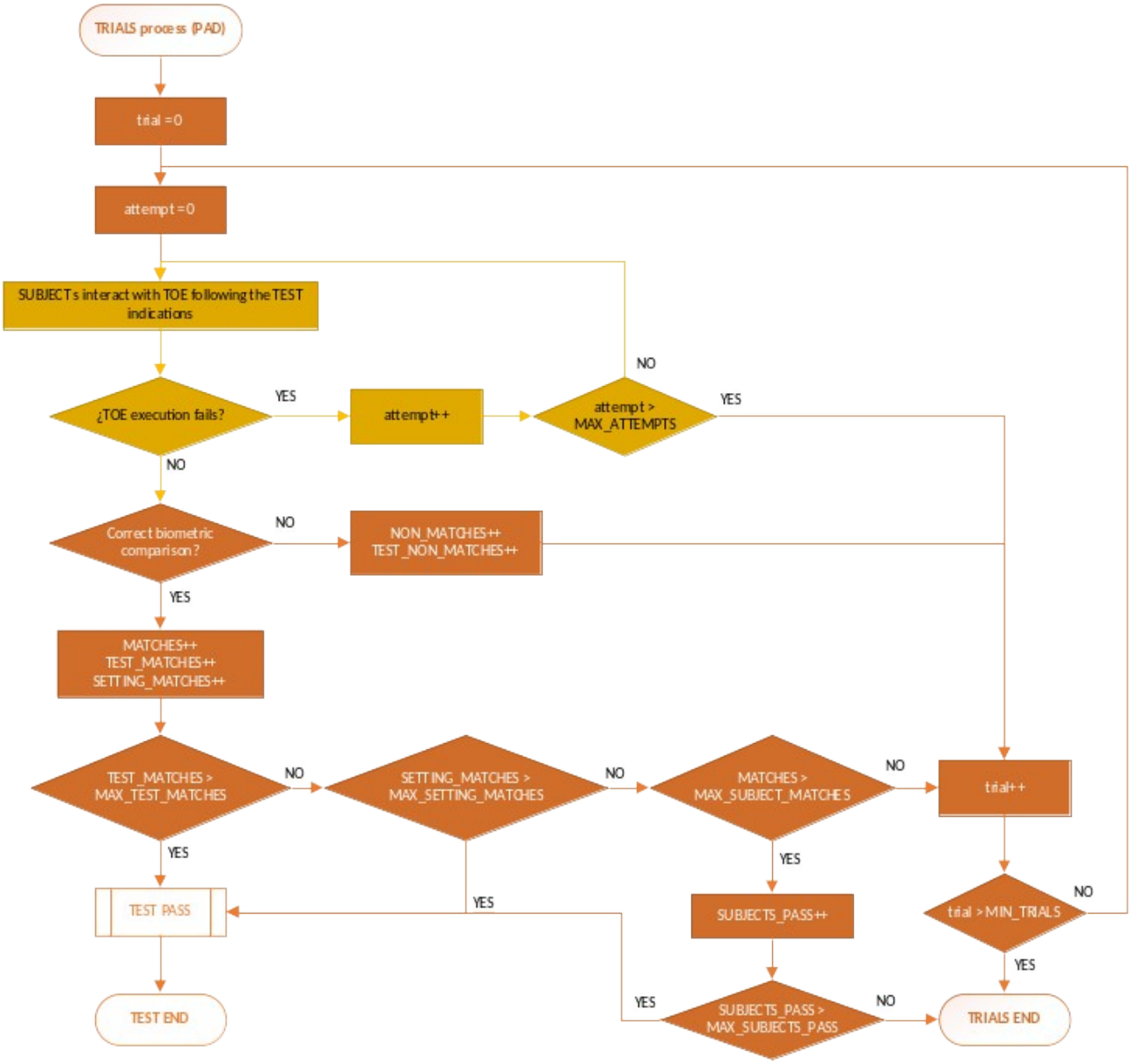

Figure 8 – Flowchart for the TRIAL-level process (Phase 4)

8.5 Families of tests in Phase 4

Many different tests can be defined for Phase 4, and this will be detailed in parts 4 or beyond, for each biometric mode.

As a general rule, tests in Phase 4 can be grouped in the following set of families:

— Zero-Effort attacks: regular use of the TOE with the intention to perform an attack.

— Enrolment-based attacks: attacking the enrolment phase, with the intent to generate a biometric reference that could be used for an easy attack in a later recognition process (e.g., enrolling the biometric features of a different person, morphing the biometric features of two persons to allow a correct recognition for both persons, etc.)

— Attacks during recognition process: attacking the TOE once the user has been correctly enrolled in the system, as to be able to impersonate such user.

NOTE For the execution of these tests, it may be useful to consider the toolboxes developed by BIO-iTC group. Examples of these toolboxes are (BIO-iTC, Fingerprint toolbox, n.d.) for fingerprint or (BIO-iTC, Face toolbox, n.d.) for face recognition. An introduction to the toolboxes is also found in (BIO-iTC, Biometric PAD Toolbox Overview, n.d.).

9 Additional methodology when evaluating machine-learning-based (ML-based) biometric products

9.1 General requirements

Some biometric products are developed with subsystems that use Machine Learning (ML) methods (e.g., Neural Networks, Markov Models, Gaussian Mixture Models, etc.). ML methods are based on the use of a set of data to train the network so that it can behave the way is expected, i.e., to get the network to learn the solution to the given problem.

ML methods can be used in the **biometric functionality subsystem** (3.1.1), for one or several tasks, such as quality checking, pre-processing or biometric comparison. Depending on how ML is used in the biometric functionality subsystem, the following classification is relevant for the evaluation of the biometric product:

— Static biometric functionality. In this case, the ML has been used during product development, and the parameters of the biometric functionality subsystems are not changed during the whole life of the biometric product. Therefore, if the biometric product has gone through an evaluation, the evaluation results can be considered valid during the whole life of the biometric product.

— Continual improvement. In this case, the biometric product can be updated, and during that update, the biometric functionality subsystem will be re-trained, changing its functionality, and therefore, requiring also an updated evaluation.

— Continuous learning. In this case, the biometric functionality subsystem goes through incremental update continuously (e.g., after any single biometric comparison). Every time the biometric functionality subsystem is updated, the evaluation results may become invalid.

For all biometric products that use some ML method, the methodology explained in this document applies. In addition to this methodology, an additional requirement is established: the data set used for training shall be different to the one used in the evaluation.

In the case of those products under the category of “static biometric functionality” no further requirement is established. But for the other two categories, further requirements are established in the following two subclauses.

9.2 Continual improvement

The products under the category of "continual improvement" suffer, from time to time, an update that will mean the retraining of the biometric functionality subsystem. Due to this retraining, the performance of the product may be impacted. Therefore, after such retraining, the product has to be re-evaluated, as if a new product is meant. In order to determine that such retraining did not have a negative impact on the system, the performance rates shall show results in half of the same order of magnitude, or even better, compared to the last evaluation carried out before the update.

EXAMPLE If before the update the biometric product shows FNMR=0,03 for an FMR=10^{-5}, after the update the product shall provide FNMR <= 0,8 for the same FMR.

It is also important to keep track of the product versions and the evaluation results, for allowing auditing processes. Last, but not least, the biometric functionality subsystem shall not be impacted by the data used for the re-evaluation. Therefore, the following process will be followed after the biometric product has been updated:

1. Create a backup of the biometric functionality subsystem
2. Execute the tests according to the relevant application profile
3. Store the backup copy of the biometric functionality subsystem, together with the evaluation results
4. Recover the backup copy of the biometric functionality subsystem, as to avoid any change that the evaluation may have caused to the biometric functionality subsystem.

9.3 Continuous learning

9.3.1 Introduction

A biometric product under the category of "continuous learning" is an extreme case of a "continuous improvement" product. In this case, the updates are every time the product is being used. Therefore, theoretically, the product should be evaluated every single time. Obviously, this is not viable, and a periodic evaluation shall be established.

In order to define the evaluation methodology for this kind of products, the following subclauses define the parameters needed.

9.3.2 Period between evaluations

When the biometric product is being designed and developed, the PM has to define an **initial period** for re-evaluating the system. Such initial period is an estimation about how long it will take for the biometric functionality subsystem to change its performance significantly. In case the PM does not define a value for such initial period, it will be taken as 4 natural days (i.e., 96 hours)

After the initial evaluation, the re-evaluations will take place at the frequency given by such initial period, establishing it as the **current period**.

At each re-evaluation, the evaluation results are compared with the ones obtained in the **previous evaluation**. If those results are equal or better than the previous ones, then, the current period will be multiplied by 2. In other case, the current period is divided by 2.

In any case, the current period cannot be lower than 24 hours.

The parameters used for these decisions are both FTA and FNMR, at an FMR defined by the application profile.

EDITOR'S NOTE: We may think about any other parameter for Phase 4 tests, in case they are applicable (see following subclauses).

9.3.3 Evaluation time lapse and infrastructure

As the evaluation may take place even daily, the only viable way to perform such evaluation is by remote means. Therefore, it shall be established a remote connection between the TL and the biometric product, so that the TL may request a re-evaluation at any desired time, following the rules given in this document.

Such remote connection shall include the following rules:

— The biometric product shall not cache or store the biometric samples sent by the TL

— All communication between the biometric product and the TL shall be protected by state-of-the-art security mechanisms, in order to provide confidentiality, authenticity and integrity of all data exchanged.

— The communication shall be done using ISO/IEC 30108 series, as to ensure interoperability among TLs and biometric products.

As it will be a remote evaluation, it shall be a technology evaluation, i.e., using a data set that will be provided by the TL to the biometric product. Such data set shall be defined by the TL, in order to comply with the following requirements:

— It has to be representative of the target population intended for the deployment of the product.

— It has to be as realistic as possible, approaching the real biometric samples received by the biometric functionality subsystem.

— The size of the data set shall be such that it will allow the re-evaluation to take a maximum of 1 hour.

— The data set has to contain some samples that will evaluate the behaviour of the biometric product against some of the Presentation Attacks (PAs)

9.3.4 Evaluation procedure

Following the specifications given in clause 9.2, the evaluation methodology for this kind of biometric products is the following:

1. Before starting the re-evaluation, a complete evaluation of the biometric product shall be carried out, according to the applicable application profile.
    a. Create a backup of the biometric functionality subsystem
    b. Execute the tests according to the relevant application profile
    c. Store the backup copy of the biometric functionality subsystem, together with the evaluation results
    d. Recover the backup copy of the biometric functionality subsystem, as to avoid any change that the evaluation may have caused to the biometric functionality subsystem.

2. If the biometric product passes the evaluation criteria provided by the application profile, after the recovery of the backup copy a first re-evaluation is performed to establish the **initial evaluation results**.
3. The **current period** is established following clause 9.3.2, and the **current evaluation results** is established as the ones obtained as initial evaluation results.
4. Whenever it corresponds, a re-evaluation is performed, using the same procedure given in clause 9.2, including backing-up and restoring of the biometric functionality subsystem.
5. If the evaluation results are worse than half an order of magnitude from the initial evaluation results, the TL will launch an alarm indicating a misbehaviour of the biometric product, applying the applicable rules established by the certification scheme.
6. Execute the tests according to the relevant application profile
7. Store the backup copy of the biometric functionality subsystem, together with the evaluation results
8. Recover the backup copy of the biometric functionality subsystem, as to avoid any change that the evaluation may have caused to the biometric functionality subsystem.

The parameters used for these decisions are both FTA and FNMR, at an FMR defined by the application profile.

EDITOR'S NOTE: We may think about any other parameter for Phase 4 tests, in case they are applicable.

EXAMPLE If the initial evaluation results show FNMR=0,03 for an FMR=10^{-5}, if the current evaluation results show an FNMR > 0,8 for the same FMR, the TL will launch an alarm indicating a misbehaviour of the biometric product.

Bibliography

BIO-iTC. (n.d.). *Biometric PAD Toolbox Overview*. (BIO-iTC) Retrieved 12 01, 2023, from https://biometricitc.github.io/#_current_documents

BIO-iTC. (n.d.). *Face toolbox*. (BIO-iTC) Retrieved 12 01, 2023, from https://github.com/biometricITC/Face-Toolbox

BIO-iTC. (n.d.). *Fingerprint toolbox*. (BIO-iTC) Retrieved 12 01, 2023, from https://github.com/biometricITC/Fingerprint-Toolbox

Comission, E. (n.d.). *The EU Cybersecurity Act*. Retrieved 12 13, 2023, from https://digital-strategy.ec.europa.eu/en/policies/cybersecurity-act

ISO/IEC_JTC1/SC37_WG3. (n.d.). ISO/IEC 19794 (all parts), Information technology — Biometric data interchange formats.

Laboratory, B. S. (n.d.). *Fingerprint Generation*. (DISI - University of Bologna) Retrieved 11 19, 2023, from http://biolab.csr.unibo.it/research.asp?organize=Activities&select=&selObj=12&pathSubj=111%7C%7C12&Req=&

Union, E. (2019, 04 27). *Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 52*. (EUR-Lex) Retrieved 11 20, 2023, from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32019R0881