BEAT

Biometrics Evaluation and Testing

http://www.beat-eu.org/

Funded under the 7th FP (Seventh Framework Programme)

Theme SEC-2011.5.1-1

[Evaluation of identification technologies, including Biometrics]

D6.5: Towards the Common Criteria evaluations of biometric systems

Due date: 02/29/2016 **Submission date:** 03/11/2016
Project start date: 01/03/2012 **Duration:** 48 months
WP Manager: Nils Tekampe **Revision:** 1.0

Author(s): N. Tekampe (TUViT), A. Merle, (CEA), J. Bringer (Morpho), M. Gomez-Barrero (UAM), J. Fierrez (UAM), J. Galbally (UAM)

Project funded by the European Commission in the 7th Framework Programme (2008-2010)		
Dissemination Level		
PU	Public	YES
RE	Restricted to a group specified by the consortium (includes Commission Services)	No
CO	Confidential, only for members of the consortium (includes Commission Services)	No

D6.5: Towards the Common Criteria evaluations of biometric systems

Abstract:

This document provides guidance for the evaluation of biometric systems and devices according to the Common Criteria for Information Security Evaluation. It is primarily targeted to the evaluator of a biometric system but also contains relevant information for the vendor of the system and the certifier.

Contents

1 Introduction

Biometric systems today are widely used in areas that require a certain level of security and assurance about the used technology. Classical examples for such applications include access control systems to high security areas (like power plants or data centres) and border control systems. Those areas usually require a high degree of assurance in that the used technology is operating as specified and as needed to obtain a secure system. In order to achieve this assurance, independent evaluations and certifications are carried out for the important components of a system or the whole system. The de facto standard for evaluations and certification of components and systems in the area of Information Security are the Common Criteria for Information Security evaluation ([7]).

While most of the relevant components used in important areas have been independently evaluated and certified, this is often not the case for the biometric systems.

The reasons for this lack of assurance are diverse but one important aspect is that up to today there is no comprehensive guide existing for the evaluation of biometric technology. Evaluations of biometric components in the past have shown that the Common Criteria are in principle applicable to biometric technology. However, some intrinsic aspects of the biometric technology require interpretation of the criteria. Without a comprehensive and accepted guidance those interpretations will have to be taken in the course of each evaluation. This leads to a lack of comparability of evaluations taken by different evaluation laboratories and also leads to a high degree of uncertainty for the developer.

The smart card community that is using the Common Criteria extensively has shown that a comprehensive set of guidance documents (e.g. [4]) and an active community is beneficial for all parties in this area.

This document aims to provide the evaluator of a biometric system with guidance on the intrinsic characteristics of the biometric technology and how they should be treated with during evaluation. It aims to provide a comprehensive guide on the evaluation of biometric components and systems according to the Common Criteria. This document is also directed to developers of biometric systems who aim to undergo an evaluation according to Common Criteria. As some of the requirements from the criteria are extended, the developer should be aware of this guidance before starting an evaluation.

In order to achieve this, the document is structured as follows:

Chapter 2 provides an overview over the general approach and provides guidance on the assurance classes ADV, AGD, ALC, APE, and ATE.

Chapter 3 provides detailed guidance on the test of a biometric system in the context of the assurance class ATE

Chapter 4 provides detailed guidance on the vulnerability assessment of a biometric system in the context of the assurance class AVA

Please note that this document presupposes a **detailed** knowledge of the scheme of the Common Criteria and will not introduce its concept or vocabulary.

2 General approach

2.1 Introduction

The Common Criteria for Information Security Evaluations are the de facto standard when it comes to to independent security evaluations in the area of IT security. They have been developed independently of any concrete technology (though they have been influenced significantly by some) and as such they claim to be usable for any IT security system.

Experiences of the last years have shown that the Common Criteria are also applicable to biometric systems. Due to some intrinsic characteristics of the biometric technology however, there is a need for additional guidance and interpretation for the security evaluation that is provided in form of this document.

As a general statement for the rest of this document serves the following adopted paragraph from [8]

> **In general, the assurance requirements from part III of [7] to establish that the system's functional requirements and specifications are realised in its development and implementation are considered to be the same for biometrics as for any IT security system or component. In this context, all classes, families and components of assurance are applicable to biometric systems.**

For some aspects during the evaluation special attention will need to be paid to the intrinsic characteristics of the biometric technology. This is specifically the case for the area of testing and vulnerability assessment. Also for the rest of the assurance classes some dedicated aspects need to be considered.

Therefore, the following paragraphs of this chapter will provide the evaluator with additional information about the way, biometric technology shall treated with during an evaluation.

2.2 General guidance

The following paragraphs provide the evaluator of a biometric system with some additional information that should be considered during the evaluation. The guidance is structured after the assurance classes from part III of [7] as listed in the following table.

Assurance Class	Additional Guidance
ASE and APE: Security Target and Protection Profile evaluation	see chapter 2.2.1
ADV: Development	see chapter 2.2.4
AGD: Guidance Documents	see chapter 2.2.5
ALC: Life-Cycle-Support	see chapter 2.2.6
ATE: Tests	see chapter 3
AVA: Vulnerability Assessment	see chapter 4

Table 1: Overview of Assurance Classes and their additional guidance

Please note that table 1 contains a more detailed overview over all assurance families that guidance is provided for.

2.2.1 Security Target and Protection Profile evaluation

In general the requirements for the evaluation of Security Targets and Protection Profiles should be applicable to biometric systems. The following aspects shall be specifically considered by the evaluator:

- The biometric system shall be clearly identified as a biometric system and its biometric functionality shall be clearly described in the ST Introduction.
- A clear demarcation of the TOE is essential. It is specifically important to identify the functionality that is part of the TOE and the functionality which is presupposed/assumed (e.g. the enrolment process).
- The working points of the biometric systems (the working point comprises concrete values for all settings that influence the performance of the biometric system) and their implications to the functionality shall be identified in the Security Target.
- It is beneficial to utilize existing standards when describing the biometric system.
- To avoid misunderstandings a continuous vocabulary should be used. The use of [1] may be beneficial in this context.
- Within the Security Problem Definition, any aspects in which the biometric system relies on specific characteristics of its environment shall be clearly stated in a way that an end user can understand them. This is of specific importance as the Security Target is the starting point for a user who is interested in a certified product. It is important to understand the concrete application case of the biometric system and the assumed environment. Aspects of the environment (e.g. light, humidity) may have a direct impact on the performance of the biometric system. It is essential for the operator of the biometric system to understand those relations. Also, biometric systems often need a dedicated supervision in order to avoid certain kind of attacks.

For the description of the functionality of the biometric system please refer to 2.2.3

2.2.2 The definition of the TOE

As outlined before, a clear demarcation of the TOE is an essential prerequisite for each Common Criteria evaluation. In the area of biometric systems evaluations, the decision about the nature of the TOE also leads to further implications that reach into the areas of ATE and AVA. As an example the question, how the relevant performance rates of a biometric system as a TOE can be tested is mainly determined by the question what actually the TOE is. In case the TOE is only made up by a biometric algorithm, it is obvious that such a test is to be performed utilizing a test database and a technical test setup. If the TOE comprises a complete biometric system including a biometric capture subsystem (with a hardware sensor), it is very likely that relevant parts of the test have to be carried out in form of a laboratory test with a test population of real people. Also the question, whether presentation attack (also commonly called spoofing) scenarios will need to be considered during the vulnerability analysis, is highly depending on the type of the TOE. For a pure algorithm it may be a suitable solution to pose an assumption about the intended environment that the environment will block those attacks. For a TOE that comprises a complete system including the sensor, such an assumption may be more difficult to take. The following list identifies a set of typical types of TOEs from the biometric world and gives some guidance about their special aspects that will need consideration. Please note that this list does not claim to be complete:

A software only TOE As a general rule it is desirable to evaluate a biometric system that covers the complete biometric functionality (from enrolment to verification) and including all relevant parts that are needed for it. However, in certain cases, it may also be useful to define a software only TOE that only comprises an algorithm for comparison or presentation attack detection (for anti-spoofing). This is of specified interest in cases of composed systems in which one developer only provides the algorithm. In such a case it may be useful to evaluate the security characteristics of the algorithm under appropriate assumptions about its environment first. Afterwards, the algorithm can be integrated into a wider system scope and a new evaluation of the complete system may reuse the results of the evaluation of the algorithm. Another field in which a pure software TOE may be desirable is the smart card world. A comparison-on-card (or match-on-card) system for example would usually only comprise the software for comparison being dedicated to work on a certified hardware platform.

A biometric capture device only In former times it has often been said that a pure biometric capture (sensor) device shall not form the TOE for a dedicated evaluation. The reason for this has always been the opinion that a capture device does usually not contain a sufficient amount of security features. However, in the context of presentation attacks against biometric systems this has changed. It is often the case

that direct (spoofing) attacks against a biometric system are countered by the capture device as the countermeasures include mechanisms in hardware. In this context it can also make sense to define the capture device of a biometric system being a TOE and being responsible for presentation attack detection.

A complete system including the sensor The typical and most desirable case in a Common Criteria evaluation is and should be the case that a complete biometric system is defined as the TOE comprising all the relevant security characteristics.

2.2.3 The functional view to a biometric system

When a biometric system should be evaluated and certified it is clearly desirable to refer to a listed Protection Profile. The following table provides an overview of the existing Protection Profiles by May 2013.

Protection Profile	Revision	Shortcut	Date	Certification ID
Common Criteria Protection Profile Biometric Verification Mechanisms	1.04	n/a	08/17/05	BSI-PP-0016-2005
Biometric Verification Mechanisms Protection Profile	1.3	BVMPP	08/07/08	BSI-CC-PP-0043-2008
Fingerprint Spoof Detection Protection Profile	1.8	FSDPP	11/23/09	BSI-CC-PP-0063-2010
Fingerprint Spoof Detection Protection Profile based on Organisational Security Policies	1.7	FSDPP-OSP	11/27/09	BSI-CC-PP-0062-2010

Table 2: Overview of Protection Profiles referenced to Biometrics

However, there may be good reasons to use a proprietary Security Target for the evaluation of a biometric system. In this case the aforementioned Protection Profiles should be taken into account when modelling the biometric functionality in terms of SFRs from part II of [7]. In a case, where a proprietary ST should be used for an evaluation also standards from the ISO world such as [10] or [11] can be useful in order to come to a sound and comprehensive description of the TOE.

2.2.4 Class ADV: Development

The requirements of the Development class provide information about the TOE. The knowledge obtained by this information is used as the basis for conducting vulnerability analysis and testing.

The following aspects should be specifically considered during the evaluation of all deliverables of the ADV class:

- It is of specific importance that the biometric functionality of the TOE and its scientific background is explained in sufficient detail so that the evaluator is able to plan their activities in the area of testing and vulnerability assessment. The evaluator needs to understand the way the biometric system analyses the biometric characteristic of the user and the way the biometric data are processed.
- The design of the TOE should be developed in a way that the biometric functionality is grouped into separate subsystems and modules. As a description of the interfaces of those subsystems and modules it may be useful to utilize existing standards (such as the 19794-series of ISO/IEC SC 37).
- The working points of the biometric systems (the working point comprises concrete values for all settings that influence the performance of the biometric system) and their implications to the functionality should be explained in detail.

2.2.5 Class AGD: Guidance Documents

The guidance documents class provides the requirements for guidance documentation for all user roles. For the secure preparation and operation of the TOE it is necessary to describe all relevant aspects for the secure handling of the TOE. The class also addresses the possibility of unintended incorrect configuration or handling of the TOE.

In the course of the evaluation of the deliverables for the AGD class the evaluator shall pay special attention to the following aspects:

- The Guidance documents need to explain all settings and parameters that influence the performance of the biometric system and their limitations.
- For the administrator it is specifically important to explain all environmental characteristics that the TOE relies on for its correct operation.
- It is important that the Guidance documents explain the functionality of the biometric system and how it should be used to all relevant roles including the end user. It may be the case that a biometric system is under evaluation that is not intended to be used by an end user directly but that has to be integrated within a larger system. In this case the integrating party shall be considered in the Guidance Documents as well.
- If the TOE relies on its environment to provide certain functionality (e.g. an enrolment), it is important to explain all the requirements that this functionality needs to fulfil in detail.

2.2.6 Class ALC: Life-Cycle-Support

Life-cycle support is an aspect of establishing discipline and control in the processes of refinement of the TOE during its development and maintenance. Confidence in the

correspondence between the TOE security requirements and the TOE is greater if security analysis and the production of the evidence are done on a regular basis as an integral part of the development and maintenance activities. ([7]) Some mechanisms for presentation attack detection work based on signature files that describe the specific characteristics of a presentation attack instrument (PAI). As soon as a characteristic is detected that matches a signature file, the system will assume that this is a presentation attack. Such signature files make a system more flexible as they can usually be updated easier than the rest of the product. As soon as a new PAI type is known, it is possible to add a new signature file. Within the ALC class it is essential to have such signature files under version control as they are essential for the operation of the TOE.

3 Testing of biometric systems

3.1 Introduction

As outlined in chapter 2 a biometric system in the context of an evaluation according to Common Criteria is handled as any other IT-system. Also the requirements from the assurance class ATE are applied as for any other system. Biometric systems, however, present some intrinsic characteristics that have to be taken into account in their testing. In particular, one main feature of biometric systems that is not shared with other IT solutions is that biometric systems are not deterministic but probabilistic systems that have intrinsic error rates associated to their normal operation in the task of identifying users. Also the fact that direct attacks against biometric systems (i.e. presentation attacks) are well known and that some systems provide countermeasures against these attacks needs to be taken into consideration during testing.

A test of the security relevant error rates is an important aspect of every Common Criteria evaluation of a biometric system. Further, the requirements in Common Criteria ensure that also the developer of a biometric system under evaluation will have to test the error rates of the system under its normal operation mode.

Depending on the question whether mechanisms for presentation attack detection are part of the TOE or not, i.e. whether they have been modelled in form of a Security Functional Requirement (SFR) or not, then also a test of these capabilities needs to be carried out.

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

The technology specific aspects in this chapter have been developed under consideration of the requirements in ISO/IEC 19795-1 ([2]).

3.2 Biometric error rates under the normal operation scenario

The class ATE refers to the tests that have to be performed to assess the performance of the security system under its normal operation scenario (see D3.3 for further details).

In the case of a verification biometric system the normal operation scenario may be defined as follows: "a legitimate enrolled user tries to access the system as himself". In this scenario two type of access attempts may be distinguished:

Definition 3.1 (*Genuine attempt or biometric mated comparison trial*)

> *in which an individual submits his/her own biometric characteristics attempting successful verification against his/her own template*

Definition 3.2 (*Impostor attempt or biometric non-mated comparison trial*)

> *also refer to as "zero-effort" impostor attempt, in which an individual submits his/her own biometric characteristics as if he/she were attempting successful verification against his/her own template, but the comparison is made against the template of a different user*

According to these two type of access attempts two different type of decision error rates have to be evaluated on verification biometric systems under the normal operation scenario (the FAR and the FRR) which are intimately related to a respective comparison error rate (the FMR and the FNMR):

Definition 3.3 (*False Non-Match Rate, FNMR*)

> *proportion of genuine access attempts falsely declared to not match the compared self template.*

Definition 3.4 (*False Reject Rate, FRR*)

> *proportion of verification transactions with rightful claims of identity that are incorrectly rejected.*

Definition 3.5 (*False Match Rate, FMR*)

> *proportion of impostor access attempts falsely declared to match the compared non-self template.*

Definition 3.6 (*False Accept Rate, FAR*)

> *proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed.*

The difference between an access attempt and a verification transaction is that FNMR and FMR concern only the comparison process while the FRR and FAR are system errors (that also include other type of errors such as the Failure to Enrol or the Failure to Acquire, defined below). Therefore, in practice, in the evaluation of a whole biometric verification *system* the pair FAR/FRR should be assessed.

The FAR/FRR of a system depend directly on the operating point selected for a given application. The operating point is defined as the value of the decision threshold where the system distinguishes between genuine and impostor access attempts.

Although very similar, as mentioned above, the FAR/FRR of a verification system also include other type of errors different from the FMR/FNMR, which are related to the sample acquisition process, namely:

Definition 3.7 (*Failure to Enrol Rate, FtER*)

is the expected proportion of the population for whom the system is unable to generate, in the enrolment phase, repeatable templates. This will include those unable to present the required biometric feature, those unable to produce an image of sufficient quality at enrolment, and those who cannot reliably match their template in attempts to confirm the enrolment is usable. The failure to enrol rate will depend on the enrolment policy. For example in the case of failure, enrolment might be re-attempted at a later date.

Definition 3.8 (*Failure to Acquire Rate, FtAR*)

is the expected proportion of transactions for which the system is unable, in the recognition phase, to capture or locate a biometric test sample of sufficient quality. The failure to acquire rate may depend on adjustable thresholds for biometric quality.

It should be noted that further error rates are existing for biometric systems that may become relevant for dedicated scenarios. [2] gives a complete overview over all possibly relevant error rates.

From these type of errors it will ultimately be the evaluator responsibility to decide which are relevant in the framework of a particular evaluation.

3.3 The role of presentation attack detection

Presentation attack (or spoof) detection capabilities in biometric systems are twofold when it comes to a Common Criteria evaluation. On the one hand, it is clear that presentation attacks against biometric systems need to be considered during each and every evaluation of a biometric system as they represent an obvious attack path to the TOE. On the other hand this does not necessarily mean that each certified biometric system has to provide presentation attack detection (PAD) capabilities.

However, in the context of the requirements in ATE it shall be clearly mentioned:

Requirement 1:
If a biometric system under evaluation implements functionality for PAD, this functionality shall be considered being TSF and shall be modelled in the Security Target in form of SFRs

This leads to the following situation:

1. If the developer of the systems claims to provide mechanisms against presentation attacks, those mechanisms shall be tested (in the course of the activities of the ATE class)

2. An any case, presentation attacks will be considered during the vulnerability analysis

The activities around PAD capabilities in the assurance class ATE focus on the question whether the provided mechanisms work as specified. It does not fall into the scope of the ATE class to develop approaches for an active circumvention of the mechanism. As usual, the question whether the mechanisms can be circumvented will be handled in AVA. Please refer to 4 for more details.

Specific guidance on tests for PAD capabilities of a biometric system can be found in 3.8

3.4 Specific requirements on assurance components

The Common Criteria knows the following elements of assurance components that require the evaluator to conduct their own test of the TOE. Namely:

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 execute a sample of tests in the test documentation to verify the developer test results.

Beside an EAL 1 evaluation or an evaluation under use of a dedicated assurance package all evaluations utilize both requirements. For a biometric system under evaluation it is essential for the evaluator to repeat the performance tests of the developer. Therefore, the

following requirements holds:

Requirement 2:
Every evaluation of a biometric system shall include the assurance component ATE_IND.2. If an EAL 1 evaluation or an evaluation according to a dedicated assurance package is performed, the assurance package shall be augmented by this component.

The following subsections contain dedicated guidance regarding the evaluation of those two components.

3.4.1 Specific requirements on ATE_IND.1

In every CC evaluation the evaluator will derive a subset of the TSF to be tested independently in accordance with the guidance in chapter 14.6.1.4 (paragraph 1367) of [6]. 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.

3.4.2 Specific requirements on ATE_IND.2

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 in the context of ATE_IND.2. Guidance on this subject is provided in chapter 3.6

3.5 General aspects for performance testing

The correlation of the two error rates FAR/FRR defined in Sect. 3.2 can best be illustrated by the use of ROC or DET curves. According to [2] they are defined as follows:

ROC curve receiver operating characteristic curve, ROC curve plot of the rate of false positives (i.e. impostor attempts accepted) on the x-axis against the corresponding rate of true positives (i.e. genuine attempts accepted) on the y-axis plotted parametrically as a function of the decision threshold

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), using logarithmic axes for an easier visualization.

Examples of ROC and DET curves can be found in figure 1 and figure 2.

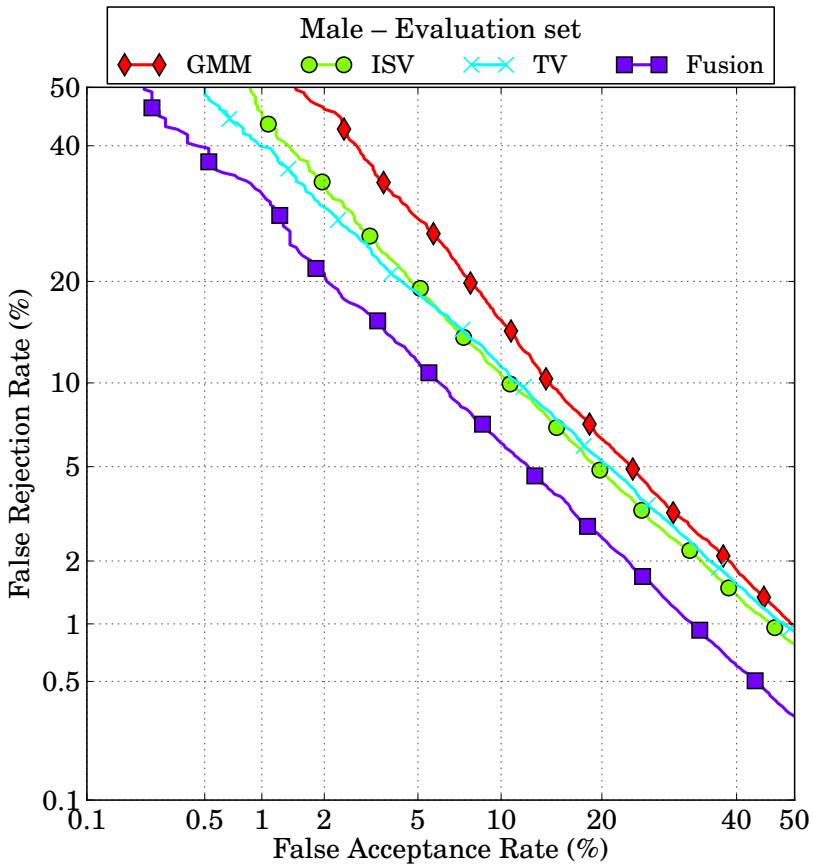

Figure 1: Example of a DET curve

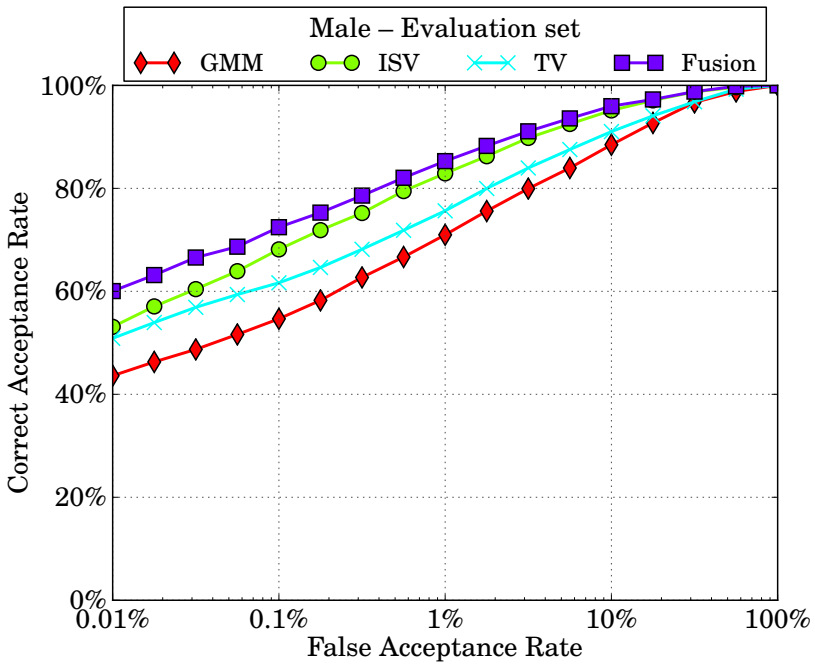

Figure 2: Example of an ROC curve

In performance testing in the normal operation scenario, ROC and DET curves are tools to illustrate the performance of the biometric system over its complete working range of decision thresholds. They are important tools to compare biometric systems and to follow the improvement of a certain biometric system over its development process.

Testing of a biometric system however is very complex and expensive. And it gets even more expensive when considering all the requirements that the Common Criteria pose on planning, execution and documentation of tests. It is therefore standard that the developer of a biometric system decides to only evaluate and certify the system at one or a very limited set of decision thresholds. The advantage of this approach is that testing will only have to be done on those decision thresholds.

It shall be clearly stated that it falls into the responsibility of the developer to decide the working point or working points of the TOE. However, it is essential that the later customer of the TOE is informed about this essential setting. Therefore, these settings shall be reported in the Security Target. Please refer to 2.2.1 for further details.

It falls into the responsibility of the evaluator to ensure that all the relevant settings of the TOE used during testing are in line with the information provided in the corresponding Security Target. Testing will then only have to be performed at those dedicated working points.

Further it should be considered that a Common Criteria evaluation is focussed on

IT-security. As an important principle it can therefore be stated that **only the security relevant error rates of a biometric system should be assessed** in the context of an evaluation.

The following paragraphs 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.

3.6 Reviewing and assessing developer tests

3.6.1 Review of developer tests

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

The developer shall implement their own test following the state of the art in performance testing.

Requirement 3:
The developer shall follow the requirements from ISO/IEC 19795-1 ([2])for their performance test. Any deviation from the requirements of the standard shall be justified in the test plan.

It is essential that the full set of information about the test is handed over to the evaluator during the evaluation. Only this way it can be ensured that the evaluator will get a complete overview over all details of the test. In this context, it is also essential that the developer will have to provide the evaluator with complete access to the used test equipment and the test data that is used for testing (e.g. in form of databases).

In the course of their analysis the evaluator shall answer and document the following questions:

- Did the developer plan, conduct and document a test in accordance with the requirements from ISO/IEC 19795-1 ([2])?
- Have all deviations from the standard been justified?
- Do the results of the test show that the biometric system under test shows a sufficient performance in its security relevant error rates?

3.6.2 Repeating a test subset

The requirements behind Assurance component ATE_IND.2 require the evaluator to repeat a sample of the tests of the developer. As outlined before, the biometric functionality of a biometric TOE is essential, so the tests of its security relevant error rates shall in any case be part of the subset to be repeated.

When repeating a test of the security relevant 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 could decide to simply repeat the actual test conduction. From a technical standpoint, the benefit of such a repetition of the technical test facilitating exactly the same test data is limited. It would only be relevant for algorithms that do not work deterministically but are self learning.

On the other hand it is a legitimate question whether the effort that the developer had to spend for test data acquisition can be re-used during the repetition of the test. This approach is also supported by the fact that the relevant results of the tests will also be backed by the results of the independent test (please see chapter 3.7).

On this basis, the evaluator is encouraged to follow the next strategy:

- Re-use the test data of the developer when repeating the test. In order to avoid a pure repetition of the test using exactly the same data the evaluator shall consider to replace a subset of the test data by their own data (which may have been obtained by any of the means described in the previous point). 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). Typically, the size of the subset to be replaced is about 20-30 %. Technically it is essential that the exchanged subset of the data is large enough to ensure that the developer could not tune their algorithm based on the database.

3.7 Conducting an independent test of security relevant error rates in the context of ATE_IND.2

ATE_IND.2 (as well as ATE_IND.3) requires the evaluator to conduct their own test of the security relevant error rates of the TOE. This chapter provides the evaluator with the corresponding guidance. The following figure summarizes the different steps that shall be performed by an evaluator when planning, performing and evaluating an independent test of the security relevant error rates of a biometric system.

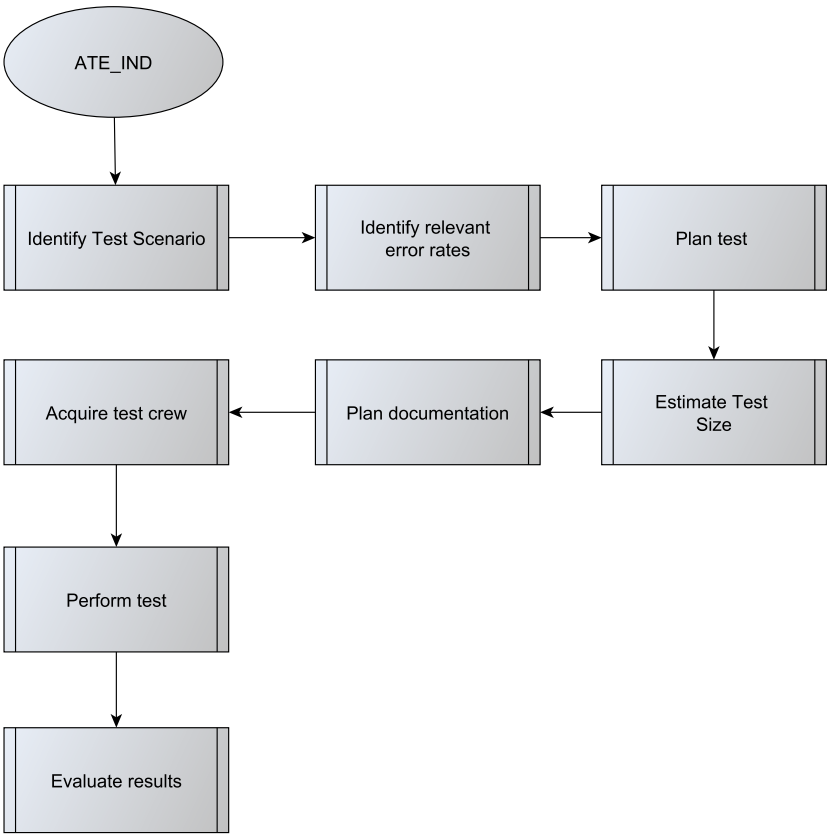

Figure 3: Process for testing

The process can be divided into the following steps that will be introduced within the following paragraphs in more detail:

Identify the relevant test scenario Various kinds of test approaches are available starting from a technological, database based test of a biometric algorithm to an evaluation of the performance of the biometric system under operation. The correct test approach highly depends on the definition of the TOE.

Identify the relevant error rates As Common Criteria focusses on the security relevant 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 has to be planned in advance.

Estimating 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 crew For the quality of results it is essential that the evaluator utilizes a test crew that is not known to the developer of the system before.

Perform Test The test is carried out under the sole control and responsibility of the evaluator.

Evaluate Results After the test has been performed, results will be evaluated and reported according to standard metrics.

Requirement 4:
The evaluator shall follow the requirements from ISO/IEC 19795-1 ([2]) for their performance tests. Any deviation from the requirements of the standard shall be justified in the test plan.

Note: In the course of the BEAT project it has been discussed whether a dedicated standard for performance tests in the context of a security evaluation may be useful. However, as such a standard is not existing as of today, 19795-1 has been used as a reference.

3.7.1 Identify the test scenario

According to [2], 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.

Operational evaluation evaluation in which the performance of a complete biometric system is determined in a specific application environment with a specific target population.

One of the very first steps for the evaluator is to identify the correct type of the evaluation for the biometric system under evaluation. As Common Criteria evaluations usually refer to an instance of a biometric product rather than to a concrete instance of an installation of a biometric system, the evaluator will usually not consider an operational test of the security relevant error rates.

The question whether a technology or scenario evaluation should be performed mainly depends on the definition of the Target of Evaluation. If a biometric algorithm is defined being the TOE is is likely that the evaluator will decide to perform a technology evaluation while a complete biometric system as TOE will more likely be tested in a scenario evaluation.

[2] 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. As outlined in [2] online tests are usually not possible for technology evaluations. Further, due to requirements regarding the repeatability and reproducibility that apply to every Common Criteria evaluation it can be stated that pure online tests (in which the images or signals will directly be discarded) shall not be used.

3.7.2 Identify relevant error rates

As mentioned before, 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 will have to consider a variety of factors to come to an answer in the course of or in preparation for an evaluation.

The following paragraphs provide an overview of the most important aspects to be taken into account in order to answer this question.

In the beginning the evaluator will have to identify the relevant error rates. As mentioned before, only the security relevant error rates are of primary interest. Those error rates determine, how likely it is that the wrong user will get recognized by the system as the legitimate one (i.e. in form of a zero effort attack). The relevant error rates depend on the operation mode of the biometric system and on the question whether a retry counter is implemented.

During an evaluation only the security relevant error rates should be assessed. For a verification system this would most likely fall back to FAR that determines the performance of the system for a particular operating point.

In order to identify all relevant error rates the evaluator shall consider all error rates that are defined in [2], the most usual ones are also introduced in Sect. 3.2, and answer two questions for each rate:

1. Is the error rate relevant for the type of biometric system.
2. Does the error rate play a role in the context/environment where the biometric system will be deployed?

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 threshold or working point of the system. A good and classical example is the False Reject Rate that is usually related to the False Accept Rate. Tuning a biometric system to show low False Accept Rate may at the same time deteriorate the False Reject Rate. Thus - as a general rule - error rates that are associated to relevant error rates shall be reported even if they do not fall into the direct scope of the evaluation. This requirement shall ensure that the biometric system stays usable and is not tuned for security reasons until it is basically unusable.

The evaluator shall check that their result of this analysis is consistent with the information that the developer provided in the Security Target.

Example on identifying relevant errors. We consider as TOE an Automated Border Control System similar to the one being developed in the ABC4EU project, considering only face verification technology, i.e., a fully automatic face verification kiosk. This is therefore an operational scenario representing an automated border control point (BCP) in an airport. This operational environment is similar to the one studied in section 2.4 of [5].

We now analyse which errors are relevant for the evaluation. A comprehensive list of possible errors (performance measures in the standard terminology) is reported in the standard ISO/IEC 19795-1:2006. In Table 3 we analyse one by one all the errors reported in that standard (most of them are described in Sect. 3.2).

As a result, the only identified error rate that must be taken into account in the evaluation is FAR. Also, as discussed in Table 3, FRR is not relevant but should be also reported to demonstrate that the system being evaluated is usable.

3.7.3 Determining maximum values for error rates

One of the most discussed questions around the evaluation of biometric systems in the context of Common Criteria is the question of the maximum value for the error rates of the system. Unfortunately, there is no absolute and simple answer to this question as it is influenced by many factors. This section provides guidance in form of a typical example on how the different factors end up in concrete values.

Example on determining maximum values for relevant errors. Similar to previous section, here we also consider as TOE an ABC Automated Border Control System similar to the one developed in the ABC4EU project, considering only face verification technology using pre-enrolled faces stored in electronic passports (eMRTD). This is therefore an operational scenario representing a border control point in an airport.

The evaluator may now investigate the state-of-the-art and in particular reports on pilots conducted by independent evaluation bodies or research groups, in addition to observing

Performance Measure	**Comment**	**Relevant? / Play a Role?**
FTE	The system uses pre-enrolment of the subjects conducted out of the scope of the TOE when the subjects receive their electronic national IDs or electronic passports (eMRTD)	No / No
FTA	The application is automated border control, where detailed knowledge of FTA or FMR is not necessary regarding security evaluation. System operation is only evaluated in the general terms provided by FAR and its corresponding FRR (see Sect. 3.2 for clarification between these 4 error rates). FTA plays a role in the context (the higher FTA the higher FRR) but it is of no relevance for the evaluation	No / Yes
FNMR	idem (in this case the higher FNMR the higher FRR, therefore FNMR plays a role but it is not relevant for the evaluation)	No / Yes
FMR	idem (in this case the higher FMR the higher FAR, therefore FMR plays an important role, but the evaluation is focused in FAR, not in FNMR)	No / Yes
FRR	This error plays an important role, as it determines the throughput of the system. False Rejections will delay significantly the process (either because the subjects try again, or because an exception is launched for manual inspection). Anyway, as happens in the discussion in the last part of the introduction to the present section, the relevant error here is only FAR, but FRR should be also reported to demonstrate that the system is usable	No / Yes
FAR	This is the most important and relevant error, as the main purpose of the system is to avoid granting access to illicit subjects	**Yes / Yes**
identification rate	n/a (the system is in verification mode)	No / No
FNIR	idem	No / No
FPIR	idem	No / No
pre-selection alg.	idem	No / No
pre-selection error	idem	No / No
penetration rate	idem	No / No
identification rank	idem	No / No

Table 3: Example on identifying relevant errors for an Automated Border Control system.

the context factors involved in the operation of the system being evaluated. As mentioned above, there is no fixed formula for computing the desired maximum error rates, but the knowledge from previous works can be a valuable help.

In the specific example here, based on an ABC kiosk, we luckily have the recent report above mentioned [5], very comprehensive in terms of factors considered (which can be consulted there), which studies in-depth the same system and operational environment being evaluated here (Test Case number 9 in that report). From the experience reported based on the pilot at Schipol Airport in Frankfurt (in which more than 6000 travellers participated), and fixing the ABC kiosk to FAR = 0.1%, then the observed FRR was around 25%. The maximum values for error rates considered in this example are then fixed as two times the errors observed in that report, i.e., FAR = 0.2% and for information FRR = 50%.

3.7.4 Plan test

Planning the test includes two important aspects. On the one hand, it is essential that the test design represents the real world scenario that shall be tested as close as possible (while still staying under laboratory conditions). On the other hand, the statistical significance behind the test has to be considered in order to report meaningful results.

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 that allows a statement about the interval that the real error rate of a system lays in with a certain confidence.

A comprehensive plan is an absolute prerequisite for a reproducible test. The test plan shall accomplish 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 should include a very detailed description of the test data that the evaluation will be performed with, this includes for instance: number of subjects, number of samples per subject, number of acquisition sessions involved, environment and external conditions of the acquisition (e.g., background, illumination, pose).
- The test plan should include a very clear protocol on how the test data is used or how test subjects shall interact with the TOE, including how the test data is divided into a train, development and test datasets and the purpose of each of the three datasets.

In particular the test dataset should be in turn divided into a set of genuine users and of impostor users (in order to generate genuine and impostor sets of scores).

- The test plan should also include a very clear protocol on how are computed the comparison scores from which the final results are extracted, that is: what users are compared and what samples of those users are used for enrolment and which for testing. In particular it should be clearly stated how the sets of genuine and impostor scores are obtained (sets of scores which later be used to derive the error rates), and the size of those sets of genuine and impostor scores.

An example of a test plan is described in A.3 of [3] as follows:

Phase	**Step**	**Activity**
Data Extraction	1	Construct three partitions • E, the first sample of each person representing an enrolment sample. • U, the second sample of each person in E, representing a user sample. • I, a sample from each person not in E.
Execution	2	Enrolment 1. Initialize supplier's enrollee data structure (EDS). 2. For each sample from E run biometric reference generator: • time the operation, store result. • if not failure to enrol append biometric reference to EDS. • Record proportion of samples that were declared unenrollable and compute failure to enrol. 3. Finalize EDS. Time this operation and store result.

	3	Sample feature extraction 1. Initialize supplier's enrollee data structure (EDS).Shuffle the N elements of U. Retain the permutation (so as to link matches back to those in E). 2. Run feature extractor on all M raw samples from U and I 3. Time each operation, store result. 4. Record proportion of samples declared unusable and compute failure to acquire. 5. Store the (non-failure to acquire) sample features.
	4	Make transaction lists 1. Make an empty list A. 2. For each sample feature from persons in U, pair it with the integer index of its match in the EDS, and add it to the list A. 3. Make an empty list B. 4. For each sample feature from persons in I, pair it with all N-1 integer indices of non-matching entries in the EDS, and add it to the list B. 5. Concatenate A and B and shuffle (randomly permute) the result, C. Retain match and nonmatch statuses.
	5	Perform full cross-comparisons 1. E, the first sample of each person representing an enrolment sample. 2. U, the second sample of each person in E, representing a user sample. 3. I, a sample from each person not in E.

Table 4: Test plan in classical performance testing

This test plan allows the test of the FMR (respectively the FAR) and the FNMR (respectively the FRR) for the biometric system under test. It is important to notice that in this particular example the tested system is already considered to be trained and that there is no threshold value (or any other parameter) to be set (typically the purpose of a development dataset). Therefore, all the users in the database are included in the test dataset. Then, one sample per user (the first sample) is enrolled to the system. The second sample is used to compare against the enrolled sample producing the user/genuine scores. Another sample from each user is compared against the enrolled samples of the *other* users to generate the impostor scores. This is the typical example of the normal operation scenario, where genuine access attempts are computed between the enrolled sample and the test sample of the same user, while impostor access attempts are computed between the enrolled sample and the test sample of different users (see Sect. 3.2).

It is essential to note that the group of all test users is divided into two test partitions in the beginning of the test plan. A set of genuine users (U) that are enrolled in the biometric system under test and a set of impostor users (I).

3.7.5 Estimate test sizes

Acquiring and handling the test crew is one of the most challenging and most expensive tasks in each test of a biometric system. According to [2], the test crew shall be as large as practically possible. However, in order to obtain statistically reliable results, there is also a minimum size of the test crew. This depends on various factors:

1. The expected error rate (the smaller the error rate the larger the test crew).
2. The required confidence interval.
3. The amount of dependencies between the various attempts. Although this factor is difficult to quantify, it should be taken into account that the statistical significance of 1,000 scores all obtained from the same user is not the same as 1,000 scores obtained from 1,000 different users.

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

As a general rule it may stated that it is always better to have many users with few samples than very few users with many samples. The higher the number of users, the lower the interdependencies between the computed scores and therefore, the higher the statistical significance of the results.

3.7.6 Plan documentation

It is essential to plan the complete documentation for the test before starting any other activities. The documentation shall follow the same principles as for any other test in the context of Common Criteria. In addition, the following aspects shall be addressed specifically:

- The exact test scenario should be carefully described.
- The relevant error rates shall be identified and their maximum acceptable values shall be identified and justified.
- The demographic characteristics of the test crew 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, users and samples per user.
- 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 are the different score sets (typically genuine and impostor) computed.

3.7.7 Acquire test crew

As already mentioned one of the key points for the evaluation of a verification biometric system is the test crew/test data that will be used in the experiments. Such data/crew can have been acquired in different scenarios: 1) by the developer for the evaluation at hand in case some specific requirements are needed (e.g., some particular illumination or background setup in the case of a face recognition system); 2) by the developer for the generic evaluation of biometric systems and not for the assessment of one particular application; 3) the developer may have obtained previously acquired data from third parties such as the multiple public biometric databases available today for evaluation purposes.

The most desirable case would be option number one, in which a different 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 third case could be sufficient if no specific contextual or external features have to be met (e.g., specific acquisition sensor). The main disadvantage of reusing already existing data is that the developer also has access to it and he may have used it to tune their system. In this situation the final results obtained would be optimistically biased as the ideal situation for an objective evaluation is to use data never "seen" before by the system, and this can only be guaranteed using private databases acquired by the evaluator.

For the acquisition of a new database or test crew, several important factors have to 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 acquisition of the test 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 control over the acquisition process.
- The acquisition of biometric test data is expensive. Therefore, it is often discussed 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 user profile (e.g., men, Asian, over 65 years of age, right-handed) the 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 and also to be able to obtain interoperability results (i.e., comparison results between enrolled and test biometric samples captured with different devices).
- It is important to design previous to the beginning of the data acquisition campaign a consistent naming convention for all the files so that each file can be tracked to a given ID. Biometric files should in no case contain the name of the real user but a generic ID code. The correspondence between ID codes and names should be kept in a separate file.
- A sufficiently large number of individuals should be enrolled in the database/crew in order to obtain statistically significant results. Such number will depend on the maximum error rates allowed for the system, the lower the error rates the larger the number of required subjects in order to achieve reliable results.
- Also, the different samples of the same user should not be captured consecutively but leaving enough time between them in order to simulate the intra-user 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 users could be also acquired. This may include for instance the gender, the age, the use of visual aids (e.g., glasses), or the

handedness. These metadata can help to further tune the performance evaluation or to reuse the biometric data in future evaluations.

- Biometric data acquisition is a very time-consuming task which is prone to many different type of human errors such as: missing samples, invalid biometric samples, errors in the naming of the files (e.g., a sample is assigned to an incorrect user), very low quality samples. Such human errors are in many cases due to tiredness of the acquisition operator and can be largely minimized if a specific semi-automatized capturing tool is developed for the campaign. Such tool can for instance automatically launch the acquisition devices (in case more than one is required) or automatically name and save the captured files.

- Biometric data is Personally Identifiable Information (PII). As such, national privacy legislations will have to be considered and followed during the acquisition process. During the acquisition the evaluator should comply with, at least, the data protection laws of his working country. This usually includes informing the acquired subjects of the use that will be given to their data and obtaining from them a signed consent form for the acquisition of those data.

3.7.8 Perform test

The previous sections defined 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 may 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 D3.3 for further details). Such standard metrics should permit to clearly evaluate the relevant errors identified for the evaluation.

3.7.9 Evaluate results

The analysis of the results should not be restricted to a mere certification of the error rates but should also include a more in-depth evaluation taking also into account in which

particular cases the system made a mistake and if those errors disclose some security problem of the system, such as for instance a significant lower performance for certain user 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.

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 may increase.

As outlined during the previous chapters, a test of the performance of a biometric system in the course of a Common Criteria evaluation is often focussed on a subset of the existing error rates. Specifically, error rates from the area of usability (such as the FRR) do often fall out of the primary scope of an evaluation. However, it should be noted that the error rates that are relevant in the context of security (such as the FAR) correlate with a dependant error rates (in this case the FRR) via the threshold of the system. Therefore, it is essential that the dependant error rate is evaluated and reported.

3.8 Testing of Presentation Attack Detection

As illustrated by the following figure, the steps performed by an evaluator when planning, performing and evaluating the presentation attack detection (PAD) capacities of a biometric system are somewhat similar to those followed when evaluating the system's accuracy. However, the stronger need for human interventions at each of these steps tends to render that evaluation more time consuming and hardly repeatable by nature.

On top of the repeatability issue, two prerequisites of any PAD system evaluation may be identified. The first one lays in the Target of Evaluation definition. We will focus on that issue in Sec. 3.8.1 but we need to underline here the importance of precisely defining the TOE and the impact it has on the whole evaluation. Another important input would be to have at hand a reference Presentation Attack Instrument (PAI) database; either PAI images acquired on the selected sensor if an algorithmic version of the remaining part of the TOE is available, or PAI materials that can typically be used. Some companies have developed presentation attack kits containing various presentation attack materials to build various PAI types. This is also the purpose of the toolbox available from the BEAT project D4.7 and D4.8 deliverables which gives a listing of available materials and fabrication techniques. Now, based on this reference PAI database, and the TOE definition, a database subset may be selected for the desired environment.

The following graph (Fig. 4) underlines the different aspects of a PAD evaluation based on these important elements.

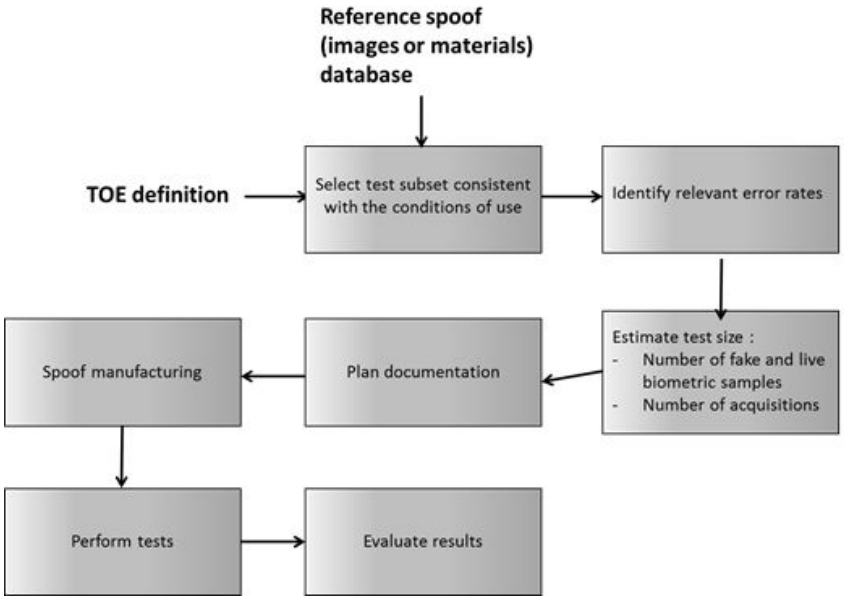

Figure 4: Different phases of PAD testing

This chapter presents the different steps to be followed during the evaluation with an emphasis on the tasks specific to the PAD subsystem evaluation. To avoid redundancies with the previous chapter, the definitions and measurements of error rates will not be repeated here. Note that, instead of using the terminology FAR and FRR to represent the success rates of presentation attack detection, the evaluator may choose to use the terminology defined in the ISO/IEC 30107 standards on Presentation Attack Detection (PAD) which define NPCER (Normal Presentation Classification Error Rate) and APCER (Attack Presentation Classification Error Rate). This avoid confusion with FAR and FRR while reporting the results from a performance testing task. Nevertheless, as this part of ATE task is not sensu stricto biometric performance testing, this is not always necessary to introduce those rates.

3.8.1 Identifying the test scenario

There are no basic types of evaluation for a presentation attack detection system. In general, we can distinguish test scenarios similar to those used when evaluating the performance rates of a biometric system:

A **"pure" technology evaluation** of the presentation attack detection software mechanism/algorithm alone such as the one performed in LiveDet. Because most PAD mechanisms are closely linked to a specific capture device (sensor), such evaluations require a learning phase for the PAD algorithm and are of limited representability for the functionality of the final product (which may use a different type of sensor and present much different results). However, a pre-existing set of PAI images may be used on all types of algorithm and ensure

the repeatability of this type of evaluation.

A **"product" technology evaluation** for which both the capture device (sensor) and PAD hardware and/or software mechanism are considered. This means performing presentation attack instrument and normal (live) biometric presentation acquisitions on an evaluation specific prototype that would return a presentation attack not detected/detected (live/fake) status on all acquired images. In that case, one will measure how many PAI were detected for a certain level of normal presentation samples rejected despite any consideration on whether the acquired samples may be further processed for features extraction or compared and thus recognized as belonging to someone enrolled in that system.

Finally, the **scenario or operational evaluations** may consider the overall system's performances in which case the extraction and comparison performances will influence the results as well. However, the evaluation will be more representative of a "real life" system, and take into account the fact that a presentation attack detection algorithm may rely on the subsequent biometric algorithms, in the remaining of biometric recognition process, to filter non-matching presentation attack instruments.

The choice of the test scenario relies on the Target of Evaluation as well as on the reproducibility expected. Several algorithms may be compared on the same dataset in a "pure" technology evaluation but it may be argued that the results are not representative of a real product's performances. The "product" technology evaluation, by considering both an acquisition sensor and the anti-spoof algorithm embedded in it limits the bias that may be introduced by the accumulation of several biometric algorithms while still allowing for a certain amount of repeatability when acquiring the same spoof and live biometric data. However, Common Criteria are dedicated to products and real systems (implementing an anti-spoofing algorithm) are to be evaluated because the certificate has to be, in the end, given to the full system. Thus, the scenario evaluation, though biased by the biometric algorithm inevitably included in the end to end system, is more representative of the targeted operational system.

On a more presentation attack oriented perspective, there is no universal golden PAI type. Indeed, depending on the acquisition scenario and type of sensor considered, the same attack may be available or not, and detected or not. Therefore, an efficient PAI (or say a fake biometric sample) is always linked to a use case and a capture device.

For example, when considering the fingerprint case, two use cases may be identified:

- **Fingerprint acquisition under operator supervision** like for passport generation. In that case, the considered presentation attack techniques need to be easily hidden to the supervisor as well as good quality PAIs (fakes) to be correctly identified by the system.
- **Fingerprint acquisition without supervision** like restricted area access control in which case all types of PAI materials may be considered by the attacker.

Whatever the TOE is, the acquisition technology used also needs to be considered, especially if one wants to use a subset of a pre-capture PAI and normal presentation acquisitions database. To date, there is no by-design presentation attack detection technology. However, the acquisition technology influences the types of materials that may be used to cheat the whole system. The main technologies for fingerprint are:

- **Optical sensors**, based on light reflection analysis, i.e. how the finger looks like. In that case the colour or the natural rigidity of the material used may be considered when creating the PAI.
- **Capacitive sensors**, based on skin electric properties, i.e. how the finger reacts conductively. In that case the attacker needs to take into account the conductivity of the PAI material used.

Finally, one must keep in mind that even though the evaluator has a wide array of techniques available, a real life attacker may not. We addressed the issue of supervised versus unsupervised acquisitions, but one also needs to consider how the attacker may obtain the target's biometric data. Hence, the target may be **cooperative**, in which case he purposely gives his identity to someone else, or **non-cooperative** if someone is trying to steal his identity by, for example, retrieving his fingerprint from a glass or taking his picture without his consent. In the former scenario, the presentation attack instrument will obviously be of much better quality (closest to the original biometric data) than in the latter.

3.8.2 Identifying relevant PAI materials and error rates

The question of the relevant error rates addressed in Sec. 3.7.2 applies to the presentation attack detection testing as well and we refer the reader to this section. It goes without saying that acquisitions of normal presentation samples (live biometric samples) need to be performed on the TOE in the same conditions as the PAI acquisitions. Indeed, the percentage of false acceptances is of interest only when related to a false reject rate: a system blocking all attacks and all live samples has no operational use. However, it seems important to focus here on presentation attack instruments acquisitions. Starting the evaluation implies that the PAI materials have been selected and it is interesting to notice that the same parameters impact both the PAI material's choice and the relevant error rates.

Usually, statistical approaches are used to measure system's biometric performances. Because we are not focusing here on a vulnerability attack but mostly assessing that the system is working properly, the number of acquisitions/tests performed may also be reduced to a bare minimum. For example, if only a subset of a predefined set of PAI (or spoof) materials is tested on a product, then one would usually consider that, among 10 acquisitions, 1 fail is OK, and 2 fails KO as the attack can then be seen as being a reproducable attack.

A stronger statistical approach may be needed, with a larger number of attempts. In that case, the evaluator shall rely on the performances metrics (in particular APCER and NPCER) defined in the ISO/IEC SC37 30107 standards on Presentation Attack Detection (PAD), in particular ISO/IEC 30107 - part 3: Presentation attack detection - testing, reporting and classification of attacks, in order to evaluate the error rates.

3.8.3 Test planning

The attacks manufacturing is one of the main issues for this type of evaluation. Indeed, when planning the tests, one also needs to take into consideration the PAIs fabrication and the reproducibility of this fabrication since some types of materials cannot be used for more than a day or two. Two elements are important here: the type of mould available to the attacker and the PAI material.

We addressed previously the problem of the mould acquisition with a non-cooperative target but aside from selecting the moulds adapted to his use case, the evaluator must decide on the spoofing materials used. As stated earlier, the material used for the PAI fabrication needs to be adapted to the use case and to the acquisition technology. Once both of these are settled, the evaluator needs to choose the PAI materials accordingly. For example, the most commonly used materials for fingerprint presentation attack instruments are:

- Paper, possibly coupled with water or alcohol to improve its conductivity.
- Glue, which may be used to extract ridges from a latent fingerprint and is therefore most likely to be used by an attacker.
- Silicon. With its different colours, and its ability to be shaped, this material is also very commonly used for fake fingers.

Depending on the capture device technology and context, gelatin, play-dough, latex, and other types of spoofing materials may be available. Other materials are also introduced for different modalities than fingerprint, as for instance HR picture of face or an iris, 3D mould of a face, etc. Again, the toolbox developed in the D4.7 and D4.8 deliverables of BEAT project lists different presentation attack instrument materials and their fabrication recipes. Therefore the evaluator may pick the presentation attack instrument types that are to be included in his reference set from this toolbox.

Once the evaluator has found volunteers for cooperative samples acquisitions and picked the relevant PAI types and corresponding materials, presentation attack instruments may finally be manufactured and tested. However, some of the materials properties have to be taken into account when preparing the test, for example, for fingerprint, PAI types based on gelatin tend to dry out and must be used within a couple of days or so while play-dough fakes need to be fabricated on the spot during the test. This means that all systems can

hardly be evaluated with exactly the same PAIs; only the types of the PAIs used can be fully replayed. Thus, it is very important to

- either prepare a high number of PAIs and to have a statistical approach measuring both the percentage of PAIs labelled normal presentation or "real" by the system and the percentage of normal presentation samples labelled presentation attack or "fake" by the system,
- or to implement procedures for the PAI creation that will ensure that PAIs made two months apart will have similar properties.

3.8.4 Estimating test size

In ATE, we propose to check that the system works as defined without actually measuring its performances on a large dataset. This means that the test size may be limited, as stated above, to, for example, 10 spoofs in each selected material.

However, this implies that the error rates of the presentation attack detection functionality cannot be measured precisely and the evaluator will need to rely and control the developer's claims, if statistically relevant rates are requested to be assessed.

3.8.5 Plan documentation

As stated in Sec. 3.7.6, it is essential to document in details:

- the selected test scenario and use case
- the relevant error rates identified
- the PAI (and moulds derived from an original biometric data) manufacturing techniques and materials as well as the manufacturing planning.

3.8.6 Perform test

Once the PAI manufacturing step is over, performing the test is quite straightforward. For each acquisition, the times and results of each step (acquisition, presentation attack detection, biometric feature extraction process, comparison process, etc.) need to be stored if available, and the number of failure at capture or enrolment as well, to allow the computation of the TOE's presentation attack detection error rates.

Note that, as explained previously, the focus of the test is not to be statistically complete but to assess that the system is working properly. In the case the evaluator chose to not rely on large number of tests, the error rates are merely reduced to counting the number of errors and success.

4 Vulnerability Analysis

This part provides detailed guidance on the vulnerability assessment of a biometric system in the context of the assurance class AVA

The contents of this chapter could be summarized in:

- Expressing the methodology to perform the AVA (Vulnerability Analysis) tasks.
- Methodology for rating the resistance of a system
- Defining Common Criteria resistance levels
- Examples of attacks and their rating

4.1 Common Criteria AVA_VAN.x tasks

The Common Criteria AVA (Vulnerability Assessment) class is based on the AVA_VAN family. It comprises 5 levels of assurance: AVA_VAN.1 to AVA_VAN.5, each level includes the lower ones and adds extra requirements.

This task introduces the idea of resistance level, rated BASIC, ENHANCED-BASIC, MODERATE, HIGH. The exact definition of these levels is left to the evaluation methodology and could be adapted to specific areas. One of the objectives of this deliverable is to propose such a definition.

Each task has some dependencies with other evaluation tasks. It is important to notice that starting at the AVA_VAN.3 level, a representation of the implementation (ADV_IMP.1) is required. This corresponds to the source code for software implementation, or very low level design for hardware (VHDL, drawing, etc.).

It has also to be noticed that AVA_VAN.1 does not include an independent vulnerability analysis (introduced from AVA_VAN.2). This means that only generic vulnerabilities (publicly known and described for example on the web, added with those identified by the developer) are taken into account by the evaluator. Starting with AVA_VAN.2 and the independent vulnerability analysis, all available knowledge, including non-published vulnerabilities and detailed knowledge of the product (the TOE) gained by the other evaluation tasks can be used to perform the task and to derive testing and attacks.

It should be noticed that biometrics systems are also hardware and software systems. Vulnerability Analysis is not dedicated only to the biometrics part; all the vulnerabilities of hardware and software have to be taken into account in targeting an evaluation level or in defining a resistance level. Description and methodology for evaluating the resistance of a system towards classical attacks are not described here because references exist (CEM for example). However, the proposed table should be used to rate successful attacks and the resistance of the TOE.

TOE resistent to attackers with attack potential of	Meets assurance components	Failure of components
No rating		AVA_VAN.1 AVA_VAN.2 AVA_VAN.3 AVA_VAN.4 AVA_VAN.5
Basic	AVA_VAN.1 AVA_VAN.2	AVA_VAN.3 AVA_VAN.4 AVA_VAN.5
Enhanced-Basic	AVA_VAN.1 AVA_VAN.2 AVA_VAN.3	AVA_VAN.4 AVA_VAN.5
Moderate	AVA_VAN.1 AVA_VAN.2 AVA_VAN.3 AVA_VAN.4	AVA_VAN.5
High	AVA_VAN.1 AVA_VAN.2 AVA_VAN.3 AVA_VAN.4 AVA_VAN.5	

Table 5: Assurance components vs. Resistance levels

4.2 Test methodology

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.

The generic approach is a "counter example" demonstration: if one attack with a lower potential than the target is found, then it is demonstrated that the TOE is not resistant to the targeted level.

So, the evaluation will focus 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 has to be done regarding the security definitions of the ST (Security Target) 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 evaluation tasks (for example knowledge of the implementation if targeting AVA_VAN.3 or higher).

The evaluation is always performed in a limited time (often defined by the National Certification Scheme) and the testing has to be limited. The objective being to find a single attack at the right level incompatible with the security specification, a "filter" could be applied to the potential tests list based on the targeted rating. From the attack specifications (what and how to do, expected results) the evaluator performs a preliminary rating. This preliminary rating prioritizes the tests done by the evaluator (the lower rating, the higher priority).

Once the tests list 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: 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? etc. This is an added value for the evaluation, justifying that, in most of the case, the maximum time allocated to the testing is used. However, it is not strictly required by the Common Criteria 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.

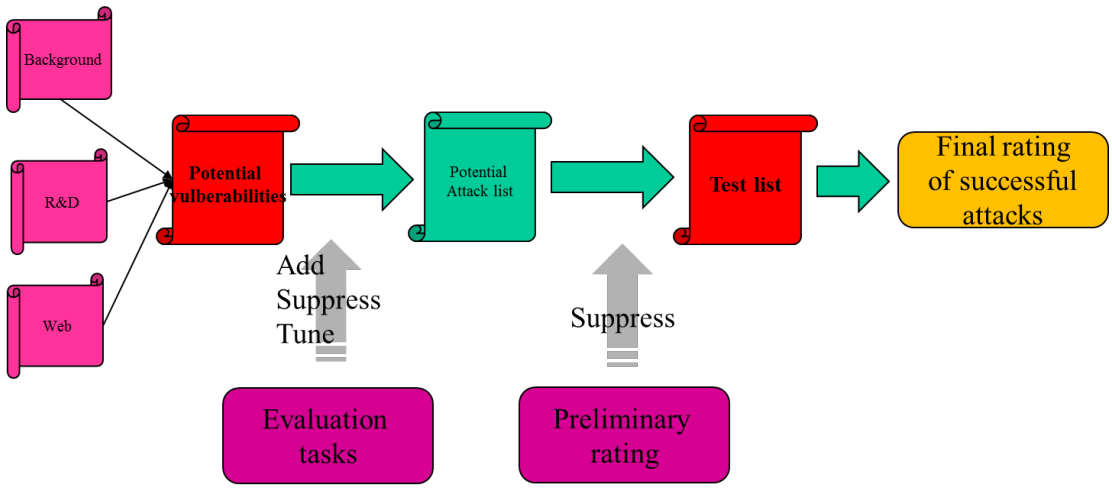

Figure 5: Testing methodology: Defining a test list

4.3 TOE for testing

Common Criteria specifies that the developer shall provide the TOE for testing and that 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, parametrisation referenced in the users 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, it should be made 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). Even if not explicitly required, it should be useful to make this equipment available to the evaluator.

4.4 Presentation attacks

For presentation attacks, there is always a part of random during the acquisition of biometrics data (for real persons or spoofs). This random is often accentuated when an aliveness detection system is added.

In addition, the chance of success when presenting a spoof is often increasing with the skills of the tester: with more time and tries, a specific strategy will probably be found to reach a higher success rate.

The suggested test methodology is the same than the one use in ATE for testing PAD:

- 10 tries are performed.
- The number of success (the spoof accepted) is counted. This is the Success Rate (SR).
- If 2 successes or more (SR >= 2) then the attack is considered as successful, if 0 or 1 success (SR < 2) then it is considered as unsuccessful

4.5 Rating an attack

4.5.1 Rating table

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2

Factor	Identification	Exploitation
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	0 (Not applicable)
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometric Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8

4.5.2 Identification/ Exploitation

Identification: corresponds to the effort required to create the attack, and to demonstrate that it can be successfully applied to the TOE (including setting up or building any necessary test equipment). The demonstration that the attack can be successfully applied needs to consider any difficulties in expanding a result shown in the laboratory to create a useful attack. One of the outputs from Identification could be a script that gives a

step-by-step description of how to carry out the attack. This script is assumed to be used in the exploitation part.

Exploitation: corresponds to achieving the attack on an instance of the TOE in its exploitation environment using the analysis and techniques defined in the identification part. Could be assumed that a different attacker carries out the exploitation, the technique (and relevant background information) could be available for the exploitation in the form of a script or set of instructions defined during the identification step. This type of script is assumed to identify the necessary equipment and, for example, mathematical techniques used in the analysis, or presentation attack methods. Furthermore, this same information may also reduce the exploitation requirement to one of time measurement, whereas the identification phase may have required reverse engineering of hardware or software information – hence the expertise requirement may be reduced.

Notes:

- For the evaluator, the work of the Identification phase has to be fully performed: developing HW and SW, creating presentation attack instruments (PAIs) if any, etc. The rating of this phase corresponds to the "real spending" in defining the attack. For the Exploitation, it is not necessary to perform the work again and the rating could correspond to an evaluation of the necessary effort for each factor.

- Exploitation consisting in applying scripts, it is expected that some factor values will be reduced from the identification phase, in particular "Elapsed Time" and "Expertise". For the same reason, the "Knowledge of the TOE" factor is not applicable in the exploitation phase (all the knowledge is scripted).

4.5.3 Elapsed Time

In the Identification phase, it corresponds to the time required to create the attack, and to demonstrate that it can be successfully applied to biometrics system (including setting up or building any necessary hardware or software equipment). The demonstration that the attack can be successfully applied needs to consider any difficulties in expanding a result shown in the laboratory to create a useful attack. One of the outputs from Identification could be, for instance, a script that gives a step-by-step description of how to carry out the attack this script is assumed to be used in the exploitation part.

Applied to presentation attacks, elapsed time in identification corresponds to the time spent to find the so called "golden fake" and to define the method to build it from for example a fingerprint (with or without the collaboration of the user). "Golden fake" is defined as a PAI that is reproducibly accepted by the TOE as being genuine.

In the exploitation phase, Elapsed Time corresponds to the time necessary to apply the "script" to a specific biometrics. For example, for a presentation attack on fingerprints, it corresponds to the time required to create a PAI from an image of a print (and not

the acquisition of this image which is taken into account in the "Access to biometrics characteristics" factor).

Potential difficulties to have an access to the TOE in exploitation environment are taken into account in the "Access to the TOE/Window of opportunity" factor.

4.5.4 Expertise

This factor refers to the level of proficiency required by the attacker and the general knowledge that he possesses, not specific of the system being attacked. A suggested rating for this metric is:

- **Layman**: no real expertise needed, any person with a regular level of education is capable of performing the attack. For example, creating a PAI in a known (published) way without specific difficulties (specific or difficult to buy materials) is considered at a "Layman" level of expertise.

- **Proficient**: some advanced knowledge in certain specific topics (biometrics) is required as well as good knowledge of the state-of-the-art of attacks. The person is capable of adapting known attack methods to his needs. For example, adapting a known PAI type (published) by the choice of specific (not published and sometimes difficult to find) materials in order to bypass a presentation attack detection mechanism and/or finding a non-evident way to present this PAI to the system can be considered at a "Proficient" level of expertise.

- **Expert**: a specific preparation in multiple areas such as pattern recognition, computer vision or optimization is needed in order to carry out the attack. The person is capable of generating his own new attacking algorithms. For example, finding a new (unpublished) way of creating a PAI type using new and specific materials (unpublished) to counter an advanced presentation attack detection mechanism, can be considered at a "Expert" level. In addition, this level can be associated with specific equipment (bespoke)

- **Multiple experts**: the attack needs the collaboration of several people with high level expertise in different fields (e.g., electronics, cryptanalysis, physics, etc.). It has to be noticed that a specific competence in biometrics is not considered as "multiple expertise". For example, building an "hill climbing" attack when the comparison score can only be accessed indirectly (for example using power or electromagnetic analysis) and when this attack requires electronic modification of the system (probing for example) can be considered at a "multi expertise" level.

Notes:

- As previously noted, Exploitation expertise is usually lower that Identification expertise. Layman or Proficient can be considered as typical value for expertise in the exploitation phase. For the same reason, the multiple expert level is excluded from the exploitation phase.

- As all the factors, higher rating would require specific justifications from the evaluator.

4.5.5 Knowledge for the TOE

This factor refers to the amount of knowledge of system required to perform the attack.

For instance, format of the acquired samples, size and resolution of acquisition systems, specific format of templates, but also specifications and implementation of countermeasures are knowledge that could be required to set up an attack.

This information could be publicly available at the website of the capture device manufacturer or protected (distributed to stakeholders under NDA or even classified inside the company).

Ratings are:

- **Public**: information which is fairly easy to obtain (e.g., on the web).

- **Restricted**: information which is only shared by the developer and organizations which are using the system, usually under a non-disclosure agreement.

- **Confidential**: information which is only available within the organization that develops the system and is in no case shared outside it.

- **Critical**: it refers to information which is only available to certain people or groups within the organization which develops the system.

Special attention should be paid in this point to possible countermeasures that may be implemented in the system and whether it is necessary or not to have knowledge of their existence in order to be successful in a given attack.

It is assumed that all the knowledge required to perform the attack is gained during the identification phase and "scripted" for the exploitation. So, this factor is not used for the exploitation phase.

4.5.6 Access to the TOE/ Window of Opportunity

This factor refers to measuring the difficulty to access the TOE either to prepare the attack (Identification phase) or to perform it on the target system.

For the Identification phase, elements that should be taken into account include the easiness to buy the same biometrics equipment (with and without countermeasures).

For exploitation phase, both technical (such known/unknown tuning) and organizational measures (presence of a guard, ability to physically modify the target, limited number of tries, etc.) should be taken into account.

The number and the level of equipment requested to build the attack is also part of the rating.

This factor is not expressed in terms of time. Proposed values are:

- **Easy:**

  There is no strong constraint for the attacker to buy TOE (reasonable price) to prepare its attack (identification phase).

  For the exploitation phase, there is no limit in the number of tries and the presentation attack is difficult to detect.

- **Medium**:

  For identification phase, specialized distribution schemes exist (not available to individuals).

  For exploitation phase, either a tuning of the attack for the final system is required (unknown parameterization of countermeasures for example), or there is a supervision of the biometrics system emitting, for example, an alert in case of numerous fail presentations.

- **Difficult**:

  For identification phase, the system is not available except for identified users and access requires compromising of one of the actors.

  For exploitation, for example PAIs must be adapted to the (unknown) specific tuning, or there is a strong supervision (for example a guard), or the system needs physical modification (for example physically accessing a hidden signal significant of the comparison score). Compromising one actor involved in the use of the system (guard, administrator, and maintenance) is often required.

4.5.7 Equipment

This factor refers to the type of equipment required to perform the attack. This includes the biometric databases used (if any). A suggested rating is:

- **Standard**: equipment which is affordable, easy to obtain and simple to operate (e.g., computer, video cameras, mobile phones, "do it yourself" material, artistic leisure materials, ...).

- **Specialized**: this refers to fairly expensive equipment, not available in standard markets and which require of some specific formation to be used (e.g., laboratory equipment, advanced printer specific materials and inks-, advanced oscilloscopes, ...).

- **Bespoke**: this refers to very expensive equipment with difficult and controlled access; for example, research printing systems with specific ink definition and flexible support adaptation. In addition, if more than one specialized equipment are required to perform different parts of the attack, this value should be used. Before using this value, it has to be carefully checked that no service is available (renting, limited time access, etc.). If such service exists, the rating has probably to be moved down to "Specialized".

4.5.8 Access to Biometrics characteristics

Common Criteria evaluations are dedicated to evaluate the intrinsic resistance of a system. Due to the potential number of attack paths (with or without the cooperation of a user for example) the evaluation does not take into account the way a real biometrics characteristic is acquired. For presentation attack detection, the vulnerability analysis is based on the hypothesis that a real "image" is available, and the rating only concerns the creation and the presentation of a presentation attack instrument (PAI).

However, it seems important to be able to compare the resistance of various systems, even based on different biometrics. In addition, getting a real "image" to build a PAI is clearly part of an attack and it seems of interest, for the final user and the pertinence of a certificate to add a factor related to this topic.

2D images can be found even without direct contact with a user (an exploration of the web and the social networks is probably sufficient); 3D images require multiple acquisitions, probably in a controlled way, without user collaboration but probably with a direct contact with him; fingerprints are left on objects the user had in hand, but need to be revealed, acquired and the corresponding images need a preprocessing; Iris images can be acquired with a high resolution camera, but with some difficulties to get a complete high quality image without user cooperation; veins are a hidden characteristic, but infra-red cameras, close to the user, can acquire images to be used.

Giving a scale or numbers has certainly an arbitrary part, and should be done by independent experts. We suggest:

Multi-modality (e.g. finger and face) or multi-instances (e.g. several fingers) systems should be rated at the upper level of the higher rate. For example, 2D images and Iris are rated *Difficult*.

Note: As explained before, rating the resistance of a system is based on rating the successful attacks and verifying that no successful attack is found at the targeted level. Some attacks do not need real biometrics data to be available, for example, attacks based

Value	Biometrics modality
Immediate	2D Face, Signature Image, Speech
Easy	Fingerprint
Moderate	Iris, 3D Face, Dynamic Signature, 3D Fingerprint
Difficult	Veins

Table 7: Proposed rating for different biometric modalities

on synthetic images or templates generation. In such a case, the value for this factor has to be set to **0**.

Note: The aforementioned explanations assume that the legitimate user(s) are non-cooperative with the attacker. If - for some specific reason - there is reason to assume that a user will give deliberately give access to their biometric characteristic, the rating value for the access to this biometric characteristic would be **0**.

4.5.9 Calculating a value

The final value is calculated by summing the value for each factor for the Identification and Exploitation phase.

4.5.10 Resistance levels

Values	Attack potential for the whole attack	TOE resistant to attackers with attack potential of	Meets assurance components	Failure of components
<10	Basic	No rating		AVA_VAN.1 AVA_VAN.2 AVA_VAN.3 AVA_VAN.4 AVA_VAN.5
10-19	Enhanced-Basic	Basic	AVA_VAN.1 AVA_VAN.2	AVA_VAN.3 AVA_VAN.4 AVA_VAN.5
20-29	Moderate	Enhanced-Basic	AVA_VAN.1 AVA_VAN.2 AVA_VAN.3	AVA_VAN.4 AVA_VAN.5

Values	Attack potential for the whole attack	TOE resistant to attackers with attack potential of	Meets assurance components	Failure of components
30-39	High	Moderate	AVA_VAN.1 AVA_VAN.2 AVA_VAN.3 AVA_VAN.4	AVA_VAN.5
>=40	Beyond high	High	AVA_VAN.1 AVA_VAN.2 AVA_VAN.3 AVA_VAN.4 AVA_VAN.5	

4.6 Examples

This chapter provides several examples, that intend to reflecting the various systems that could be evaluated (access control device for a building, an office, etc., access control to a personal device, etc.) and the "classical" attacks that could be applied.

4.6.1 Simple system without presentation attack detection

Two examples are here rated, consisting in 2D face recognition or fingerprint based systems, unattended and without any restriction to access the TOE. They only differ by the Access to Biometric Characteristics factor.

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16

Factor	Identification	Exploitation
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0 (2D face)
Easy	0 (Not applicable)	2 (Fingerprint)
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**0**	**0 or 2**
	= 0 or 2	

It is considered that:

- 1 day is enough to define the method to build a PAI type (Identification) and to generate a PAI targeting a dedicated person (exploitation). For 2D images, a simple print of a picture could be enough, for fingerprints a moulding with easy to get material (glue, silicon, latex, ...) is efficient.
- A lot of publications explain how to perform. No specific expertise is required (layman value is enough)
- No specific knowledge of the TOE is required.
- Access of the TOE does not make any problem both in identification (easy to buy without control) or in exploitation (a fingerprint PAI is easy to present: for example by "gluing" the PAI to the real finger, for 2D face, a picture is presented to the camera).

- There is no specific requirement on equipment.
- Access to biometric characteristic is rated.

The rating for the attack is: **BASIC**

The system **fails any level** of evaluation assuming that the described attack can be performed successfully.

4.6.2 Fingerprints with presentation attack detection

Let's consider a fingerprint based system with presentation attack detection.

The system is typically an access control system in an open environment. Several tries are possible but "strange" behaviour of the user would be detected.

The system includes presentation attack detection implying to find the right material for making the PAI type (glycerine, gelatin for example) and the application to a real finger is not immediate (thin film, leaving part of the skin in contact with the capture device, a print with specific ink directly on a real finger, etc.).

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		

Factor	Identification	Exploitation
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2 (Fingerprint)
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**6**	**6**
	= **12**	

It is considered that:

- Finding the right material for the PAI type and how to present it to the system is not evident and will require multiple tries. 2 weeks for identification is realistic for an example. Once defined, producing the PAI for a dedicated person and applying with the predefined method to the real TOE is immediate (1 day for exploitation).
- A lot of publications explain how to perform. However, the attacker will have to understand (and even to find) what is the principle of the presentation attack detection, and to derive a specific strategy both for creating the PAI and to apply it. A proficient level for identification is realistic, for exploitation a *layman* is enough (following a script).
- It is assumed that no specific knowledge is required. The existence of presentation attack detection is probably advertised (either by the developer or the user). With some time, the attacker (proficient level, so knowing what is offered by industrial systems), will probably find the method this detection is based on.
- *Access of the TOE*: being a security system, it is assumed that it is not possible to simply buy the system without any control, but that its distribution is controlled (for example by requiring an identification of the buyer and potentially to sign a NDA). *Moderate* level is probably adapted for the identification phase and for the exploitation phase (detection of a strange behaviour).
- There is no specific requirement on equipment.

- Access to biometric characteristic is rated.

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.

Note: The rating of the attack is close to the limit between BASIC and ENHANCED BASIC. This means that any reduction, for example less time in identification, having no control in the distribution of the system (moving from Moderate to Easy in the access of the TOE, being in a fully unsupervised environment, could reduce the rating of the attack to BASIC and then reduce the resistance of the system to "No rating".

4.6.3 Fingerprints with Advanced presentation attack detection

Let's consider a fingerprint based system with an advanced presentation attack detection.

The system is typically an access control system in an open environment. Several tries are possible but "strange" behaviour of the user would be detected.

The system includes an advanced presentation attack detection system. Advanced means that without a detailed knowledge, it is impossible to find in a reasonable time the method to make undetected PAI types (multi detection methods, needing a specific presentation strategy). In addition, no details can be found on the public domain, nor explained to buyers. It is also considered that the system is only sold to well identified users, under NDA.

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		

Factor	Identification	Exploitation
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2 (Fingerprint)
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**16**	**10**
	= **26**	

It is considered that:

- Finding the right material for the PAI type and how to present it to the system is not evident and will require multiple tries. 1 month for identification is realistic. Once defined, producing the PAI for a dedicated person and applying with the predefined method to the real TOE is immediate (1 day for exploitation).

- Presentation attack detection strategy is not described in the public domain, detailed measures are kept confidential and a specific strategy to make the PAI and present it has certainly to be invented. An *expert* level for identification is realistic, for exploitation a *proficient* is enough (following a script, but applying a complex strategy requiring good knowledge and understanding of the presentation attack detection mechanism).

- Without detailed knowledge of the presentation attack detection mechanisms, it is assumed that it is impossible in a reasonable time to create an accepted PAI. This

information is protected and a compromising is necessary. So, a *Sensitive* knowledge is required.

- Being a security system, it is assumed that it is not possible to simply buy the system without any control, but that its distribution is controlled (for example by requiring an identification of the buyer and potentially to sign a NDA). It is assumed that the protection is more efficient than in the previous example, so a *Difficult* level is probably adapted for the identification phase. For exploitation, *Access of the TOE* is rated to *Moderate*.
- There is no specific requirement on equipment.
- Access to biometric characteristic is rated.

The rating for the attack is: **MODERATE.**

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

The system is compatible with the **AVA_VAN.3** component.

Notes: A significant part of the rating is due to non-technical measures (Knowledge of the TOE, Access to the TOE). This means that if such measures are not implemented, the resistance of the TOE will have to be moved down to BASIC.

4.6.4 3D Face with presentation attack detection and try counter

Let's consider a 3D face based system with (relatively simple) presentation attack detection, but for which the detection is hard to circumvent with a chance greater than 10% and a try counter which raises an alert (and triggers some subsequent corrective actions) when more than 3 presentation attacks have been detected in a row of attempts without any successful acceptance. The system is typically an access control system in an open environment. The system includes presentation attack detection implying to find the right material for making the PAI type.

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		

Factor	Identification	Exploitation
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2
Moderate	0 (Not applicable)	4 (3D Face)
Difficult	0 (Not applicable)	8
Total	**12**	**16**
	= **28**	

It is considered that:

- Finding the right material for the PAI type and how to present it to the system is not evident and will require multiple tries. 1 month for identification is realistic for an example. Once defined, producing the PAI for a dedicated person and applying with the predefined method to the real TOE is immediate (1 day for exploitation).
- Some publications may explain how to perform. However, the attacker will have to understand (and even to find) what is the principle of the presentation attack detection, and to derive a specific strategy both for creating the PAI and to apply

it. A proficient level for identification is at least needed, for exploitation a *layman* is enough (following a script).

- It is assumed that a restricted knowledge of the TOE is required to understand what is used for PAD in case of 3D system.
- Being a security system, it is assumed that it is not possible to simply buy the system without any control, but that its distribution is controlled (for example by requiring an identification of the buyer and potentially to sign a NDA). *Moderate* level is probably adapted for the identification phase. For exploitation, *Access to the TOE* is rated as difficult due to the try counter.
- There may be specific equipment needed for producing 3D PAI during identification and exploitation.
- Access to biometric characteristic is rated.

The rating for the attack is: **MODERATE**

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

The system is compatible with the **AVA_VAN.3** component.

4.6.5 Hill climbing

Let's consider a fingerprints based system operating in a fully uncontrolled environment (for example, protecting the access to a device or an equipment). Let's also imagine that there is a way to easily connect a computer just before the matching process, enabling a program to propose templates and that there is a signal corresponding to the matching score (for example, through a debug connection). The attack will consist in proposing templates (for example starting with random locations of minutia) and to optimize the location using the matching score in a so called "hill climbing" attack.

This scenario of attack may correspond to two different objectives: One could be related to privacy leakage (to learn information on the enrolled data), a second one could be related to forging authentication (e.g. assume that the matching score or the decision cannot be modified after the matching step, the attacker would require to find a matching template to be authenticated).

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2

Factor	Identification	Exploitation
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**12**	**4**
	= **16**	

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 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.
- 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.
- A deep knowledge of the TOE is required (template formats, internal protocols, etc.). So, a *Sensitive* knowledge is required.
- The TOE operates in a fully uncontrolled environment and it is considered as easy to buy system.
- 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*.
- Access to biometrics characteristic is rated as 0 as the attack does not require the availability of real data (synthetic templates).

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.

4.6.6 Combined attack (getting matching value in an indirect way)

Let's consider the same system than the previous one, except that the matching score signal cannot directly be accessed. However, let's imagine that an indirect observation of the matching process, for example through power consumption or processing time, can give an information of the matching score (this method is widely used in smart-card evaluations and is known as Side Channel Attacks).

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16

Factor	Identification	Exploitation
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**24**	**14**
	= **38**	

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.

- More than 1 month for identification and more 1 day (less than a week) for exploitation are realistic.
- 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).
- A deep knowledge of the TOE is required (template formats, internal protocols, etc.). So, a *Sensitive* knowledge is required.
- The TOE operates in a fully uncontrolled environment and it is considered as easy to buy a system.
- 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.
- Access to biometrics characteristic is rated as 0 as the attack does not require the availability of real data (synthetic templates).

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.

4.6.7 Inverse biometrics attack

Let's consider an iris based system operating in a fully uncontrolled environment (for example, protecting the access to a device or an equipment). Let's also imagine that there is an easy way to connect a computer just before the feature extractor, enabling a program to propose synthetic samples and that there is a signal corresponding to the matching score (for example, through a debug connection). The attack will consist in proposing reconstructed synthetic samples and to optimize them using the matching score.

Note the main difference between this inverse biometrics attack and the hill-climbing attack described in section 4.6.5: the entry point of the attack, which is before the matcher for hill-climbing (i.e., after the feature extractor), and before the feature extractor for the inverse biometrics attack. The inverse biometrics attack requires therefore 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 section 4.6.5. One example of

hill-climbing attack for fingerprint biometrics appears in [12], and one example of inverse biometrics attack for hand biometrics is reported in [9].

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0
Easy	0 (Not applicable)	2
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**10**	**8**

Factor	Identification	Exploitation
	= **18**	

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.
- 2 weeks for identification and 1 day for exploitation are realistic.
- 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.
- A *restricted* level of knowledge of he TOE is required.
- 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*.
- Access to biometrics characteristic is rated as 0 as the attack does not require the availability of real data (synthetic samples).

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.

4.6.8 Dictionary attack

Let's consider an iris or face based system operating in a fully uncontrolled environment (for example, protecting the access to a device or an equipment). Let's also imagine that there is an easy way to connect a computer just before the feature extractor. However, now we don't need access to the matching score. The attack will consist in proposing real biometric samples to the system until one is accepted.

Factor	Identification	Exploitation
Elapsed Time		

Factor	Identification	Exploitation
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0 (Face)
Easy	0 (Not applicable)	2 (Iris)
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**8**	**8 or 10**
	= 16 or 18	

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.

- 2 weeks for identification and exploitation are realistic.
- 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.
- A restricted knowledge of the TOE is required.
- A specialized equipment is required (computer, connection to the system, data bases).
- 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 may be very difficult. The rating in Exploitation for Access to Biometric Characteristics reflects both facts. On the other hand, note that although it may be difficult to get a large enough iris database, once obtained it may be used for attacking different system. The rating Easy reflects that fact.

The rating for the attack is: **ENHANCED-BASIC** (face) or **MODERATE** (Iris).

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** (face) or **ENHANCED-BASIC** (Iris).

The system is compatible with the **AVA_VAN.2** (face) or **AVA_VAN.3** (Iris) component.

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 may take less than 2 weeks. We may moreover 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 may be considered as a specialized equipment.

4.6.9 Wolf Attack

Let's consider a biometric system for a given modality, operating in a fully uncontrolled environment (for example, protecting the access to a device or an equipment). Let's also assume it is a purely software-based system, for which it is thus straightforward to connect a computer just before the feature extractor. Finally, let's suppose there is a flaw in the matching algorithm such that there is a way to construct an image (not necessarily representative of a biometric capture) for which the acceptance rate is significantly higher than any randomly chosen biometric data, whatever the enrolled data are. This corresponds

to a system with a high Successful Attack Rate (SAR). The attack will consist in finding the particular (or one of) image(s) that leads to a high chance of being accepted.

Factor	Identification	Exploitation
Elapsed Time		
<= one day	0	0
<= one week	1	2
<= two weeks	2	4
<= one month	4	8
>= one month	8	16
Expertise		
Layman	0	0
Proficient	2	4
Expert	4	8
Multiple Experts	8	16
Knowledge of TOE		
Public	0	0 (Not applicable)
Restricted	2	0 (Not applicable)
Sensitive	4	0 (Not applicable)
Critical	8	0 (Not applicable)
Access to the TOE/ Window of Opportunity		
Easy	0	0
Moderate	2	4
Difficult	4	8
Equipment		
Standard	0	0
Specialized	2	4
Bespoke	4	8
Access to Biometrics Characteristics		
Immediate	0 (Not applicable)	0 (no access needed)
Easy	0 (Not applicable)	2
Moderate	0 (Not applicable)	4
Difficult	0 (Not applicable)	8
Total	**18**	**0**

Factor	Identification	Exploitation
	= **18**	

It is considered that:

- The Identification phase corresponds to find the weakness in the matching algorithm and therefore to generate an image which exploits this flaw. The Exploitation phase is just inputting the image to get access.
- More than one month for identification might be needed while exploitation is immediate.
- An expert attacker should be needed to find the flaw in the algorithm. A layman will be able to input the image once generated.
- Sensitive knowledge of the TOE is required to learn the detail of the matching algorithm.
- A specific equipment may be required to generate an image, acceptable for the feature extractor.
- Access to the biometrics characteristics is not rated as no access is needed for the attack.

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.

5 Summary

This document contains the first comprehensive evaluation methodology for TOEs from the biometric area since the BEM. In contrast to the BEM many updates have been performed. Not only that the version of the Common Criteria has been updated to the latest release, also the latest development in the area of biometrics has been considered.

For the first time, a guidance document is existing that covers the whole range of assurance classes as defined in Common Criteria. For many classes additional but simple guidance has been defined in chapter 2. This guidance allows the competent evaluator to cope with the evaluation of a biometric system.

The BEAT platform offers a framework to support testing related to the intrinsic performances of the system to evaluate.

A whole chapter has been dedicated to the area of ATE (testing) and AVA (vulnerability analysis) as these assurance classes contain the most technical questions that needed special guidance. A specific attention has been paid to presentation attacks, both from the testing methodology point of view (ATE and AVA) and from the rating of the resistance of a system. In addition, deliverables produced in other work packages (WP3 and WP4) define a basis for the testing (metrics, samples to test, methodology to produce samples).

This document has been thoroughly discussed within the course of the 4 years project BEAT where it forms the deliverable D6.5 It shows the consensus of all experts who have been involved in the BEAT project and it can be used as a basis for prototype evaluations in CC schemes.

This document is proposed as an input for further standardization activities.

References

[1] Iso jtc1 sc37 sd 2 - harmonized biometric vocabulary. Technical report, ISO.

[2] ISO/IEC 19795-1:2006 - Information technology – Biometric performance testing and reporting – Part 1: Principles and framework. Technical report.

[3] ISO/IEC 19795-2:2007 - Information technology – Biometric performance testing and reporting – Part 2: Testing methodologies for technology and and scenario evaluation. Technical report.

[4] Application of attack potential to smartcards. 1.3, 1999.

[5] Smart borders pilot project: Report on the technical conclusions of the pilot. Technical report, eu-LISA, December 2015. doi:10.2857/086263, `http://www.eulisa.europa.eu/Publications/p_reports/Pages/default.aspx?RID=28`.

[6] CCBB, editor. *Common Methodology for Information Technology Security Evaluation - Evaluation methodology*. 2009.

[7] CCBD, editor. *Common Criteria for Information Technology Security Evaluation*. 2009.

[8] D. S. A. I. S. E. P. A. K. C. A. I. S. E. P. A. V. V. A.-C. C. S. B. I. A. M. B. für Sicherheit in der Informationstechnik (BSI) Alan Richards CESG Philip Statham UK Mario Savastano Consiglio Nazionale delle Ricerche (CNR) Robert Harland Communications Security Establishment (CSE) Erin Connor EWA-Canada Dennis Weiss EWA-Canada Paul Zatychec EWA-Canada Peter Higgins Higgins and A. T. M. N. P. L. J. W. S. J. S. U. J. L. S. Oy. *Common Methodology for Information Technology Security Evaluation - Biometric Evaluation Methodology Supplement*. 2002.

[9] M. Gomez-Barrero, J. Galbally, A. Morales, M. Ferrer, J. Fierrez, and J. Ortega-Garcia. A novel hand reconstruction approach and its application to vulnerability assessment. *Information Sciences*, 268:103–121, 2014.

[10] ISO. Iso/iec 24713-1:2008 information technology - biometric profiles for interoperability and data interchange. Technical report, 2008.

[11] ISO/IEC. Iso/iec sc 37 standing document 11, part 1: Overview standards harmonization docuent. Technical report.

[12] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia. An evaluation of indirect attacks and countermeasures in fingerprint verification systems. *Pattern Recognition Letters*, 32(12):1643–1651, 2011.