INTERNATIONAL STANDARD

ISO/IEC 30107-3

First edition
2017-09

Information technology — Biometric presentation attack detection —

Part 3: Testing and reporting

Technologies de l'information — Détection d'attaque de présentation en biométrie —

Partie 3: Essais et rapports d'essai

Reference number
ISO/IEC 30107-3:2017(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

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

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

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

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

This document was prepared by Technical Committee ISO/IEC JTC 1, *Information technology*, SC 37, *Biometrics*.

A list of all parts in the ISO 30107 series can be found on the ISO website.

Introduction

The presentation of an artefact or of human characteristics to a biometric capture subsystem in a fashion intended to interfere with system policy is referred to as a presentation attack. ISO/IEC 30107 (all parts) addresses techniques for the automated detection of presentation attacks. These techniques are called presentation attack detection (PAD) mechanisms.

As is the case for biometric recognition, PAD mechanisms are subject to false positive and false negative errors. False positive errors wrongly categorize bona fide presentations as attack presentations, potentially flagging or inconveniencing legitimate users. False negative errors wrongly categorize presentation attacks (also known as attack presentations) as bona fide presentations, potentially resulting in a security breach.

Therefore, the decision to use a specific implementation of PAD will depend upon the requirements of the application and consideration of the trade-offs with respect to security, evidence strength, and efficiency.

The purpose of this document is as follows:

— to define terms related to biometric presentation attack detection testing and reporting, and

— to specify principles and methods of performance assessment of biometric presentation attack detection, including metrics.

This document is directed at vendors or test labs seeking to conduct evaluations of PAD mechanisms.

Biometric performance testing terminology, practices, and methodologies for statistical analysis have been standardized through ISO and Common Criteria. Metrics such as FAR, FRR, and FTE are widely used to characterize biometric system performance. Biometric performance testing terminology, practices, and methodologies for statistical analysis are only partially applicable to the evaluation of PAD mechanisms due to significant, fundamental differences between biometric performance testing concepts and PAD mechanism testing concepts. These differences can be categorized as follows:

a) **Statistical significance**

Biometric performance testing utilizes a statistically significant number of test subjects representative of the targeted user group. Error rates are not expected to vary significantly when adding more test subjects or using a completely different group. Generally, taking more measurements increases the accuracy of the error rates.

In PAD testing, many biometric modalities can be attacked by a large or indeterminate number of potential presentation attack instrument (PAI) species. In these cases, it is very difficult or even impossible to have a comprehensive model of all possible presentation attack instruments. Hence, it could be impossible to find a representative set of PAI species for the evaluation. Therefore, measured error rates of one set of presentation attack instruments cannot be assumed to be applicable to a different set.

PAI species present a source of systematic variation in a test. Different PAI may have significantly different error rates. Additionally, within any given PAI species, there will be random variation across instances of the PAI series. The number of presentations required for a statistically significant test will scale linearly with the number of PAI species of interest. Within each PAI species, the uncertainty associated with a PAD error rate estimate will depend on the number of artefacts tested and the number of individuals.

EXAMPLE 1 In fingerprint biometrics, many potent artefact materials are known, but any material or material mixture that can present fingerprint features to a biometric sensor is a possible candidate. Since artefact properties such as age, thickness, moisture, temperature, mixture rates, and manufacturing practices can have a significant influence on the output of the PAD mechanism, it is easy to define tens of thousands of PAI species using current materials. Hundreds of thousands of presentations would be needed for a proper statistical analysis – even then, resulting error rates could not be transferred to the next set of new materials.

b) **Comparability of test results across systems**

In biometric performance testing, application-specific error rates based on the same corpus of biometric samples can be used to compare different biometric systems or different configurations. The meaning of "better" and "worse" is generally understood.

By contrast, when using error rates to benchmark PAD mechanisms, terms such as "better" can be highly dependent on the intended application.

EXAMPLE 2 In a given testing scenario with 10 PAI species (presented 100 times), $System_1$ detects 90 % of attack presentations and $System_2$ detects 85 %. $System_1$ detects all presentations for 9 PAI species but fails to detect all presentations with the 10th PAI species. $System_2$ detects 85 % of all PAI species. Which is better? In a security analysis, $System_1$ would be worse than $System_2$, because revealing the 10th PAI species would orient an attacker such that he could use this method to defeat the capture device all the time. However, if attackers could be prevented from using the 10th PAI species, $System_1$ would be better than $System_2$, because individual rates indicate that it is possible to overcome $System_2$ with all PAI species.

c) **Cooperation**

Many biometric performance tests address applications such as access control in which subjects are cooperative. Errors due to incorrect operation are an issue of a lack of knowledge, experience or guidance rather than intent. Significant uncooperative behaviour in a group is not part of the underlying "biometric model" and would render the determined error rates almost useless for biometric performance testing.

PAD tests include subjects whose behaviour is not cooperative. Attackers will try to find and exploit any weakness of the biometric system, circumventing or manipulating its intended operation. Presentation attack types, based on the experience and knowledge of the tester, can change the success rates for an attack dramatically. Hence, it can be difficult to define testing procedures that measure error rates in a fashion representative of cooperative behaviour.

d) **Automated testing**

In biometric performance testing, it is often possible to test comparison algorithms using databases from devices or sensors of similar quality. Performance can be measured in a technology evaluation using previously collected corpuses of samples as specified in ISO/IEC 19795-1.

In PAD testing, data from the biometric sensor (e.g. digitized fingerprint images) may be insufficient to conduct evaluations. Biometric systems with PAD mechanisms often contain additional sensors to detect specific properties of a biometric characteristic. Hence, a database previously collected for a specific biometric system or configuration may not be suitable for another biometric system or configuration. Even slight changes in the hardware or software could make earlier measurements useless. It is generally impractical to store multivariate synchronized PAD signals and replay them in automated testing. Therefore, automated testing is often not an option for testing and evaluating PAD mechanisms.

e) **Quality and performance**

In biometric performance testing, performance is usually linked directly to biometric data quality. Low-quality samples generally result in higher error rates while a test with only high-quality samples will generally result in lower error rates. Hence, quality metrics are often used to improve performance (dependent on the application).

In PAD testing, even though low biometric quality can cause an artefact to be unsuccessful, there is no reason to assume a certain quality level from artefacts in general. Samples from artefacts can exhibit better quality than samples from human biometric characteristics. Absent a model of attacker skill, it seems valid (at least in a security evaluation) to assume a "worst case" scenario where the attacker always uses the best possible quality. That way, one can at least determine a guaranteed minimal detection rate for the specific test set while reducing the number of necessary tests at the same time. It is then a matter of rating the attack potential of successful artefacts (effort and expertise for the needed quality) in order to assess the security level, as is the practice in Common Criteria evaluations.

Based on the differences a) through e), the following general comments regarding error rates and metrics related to PAD mechanisms can be derived:

— In an evaluation, PAI species are analysed/rated separately.

— Attack presentation classification error rates other than 0 % for a PAI species only prove that the PAI can be successful. A different tester might achieve a higher or lower attack presentation classification error rate. Further, training to identify the relevant material and presentation parameters could increase the attack presentation classification error rate for this PAI species. The experience and knowledge of the tester, as well as the availability of the necessary resources, are significant factors in PAD testing and are taken into account when conducting comparisons or performance analysis.

— Error rates for PAD mechanisms are determined by the specific context of the given PAD mechanism, the set of PAI species, the application, the test approach, and the tester. Error rates for PAD mechanisms are not necessarily comparable across similar tests, and error rates for PAD mechanisms are not necessarily reproducible by different test laboratories.

Information technology — Biometric presentation attack detection —

Part 3:
Testing and reporting

1 Scope

This document establishes:

— principles and methods for performance assessment of presentation attack detection mechanisms;

— reporting of testing results from evaluations of presentation attack detection mechanisms;

— a classification of known attack types (in an informative annex).

Outside the scope are:

— standardization of specific PAD mechanisms;

— detailed information about countermeasures (i.e. anti-spoofing techniques), algorithms, or sensors;

— overall system-level security or vulnerability assessment.

The attacks considered in this document take place at the sensor during presentation. Any other attacks are considered outside the scope of this document.

2 Normative references

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

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

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

ISO/IEC 30107-1:2016, *Information technology — Biometric presentation attack detection — Part 1: Framework*