On the need for an Australian Identity Commissioner

Dr Ted Dunstone
Biometix
ted@biometix.com

Abstract

The following article explains the need for a government identity commissioner.

1. Reporting API: Overview

1.1. API Usage

The api has one main endpoint `/report` and accepts both post and get requests. A post request is used to create a report job. A get request is used to retrieve a report from the Performix reporting tool.
Note. Usage of the API is authorized by using a bearer token. Please contact Biometix for a valid auth token.

1.1.1. Creating a report job

The Performix reporting tool is designed to be stateless. Therefore all data and associated specification files must be provided in the form-data of the post request.

The reporting tool requires:

- a blueprint yaml file that specifies the analyses to be performed.
- a template file that specifies the format of the report as a word docx document.
- one or more data files in csv format.

For further information on how to set up the blueprint and template files please see the Blueprint Specification document.

A post to `/report` accepts a multipart/form-data content type with 3 named keys:

- blueprint - for the yaml blueprint configuration file.
- termplate - for the docx template file.
- data - for csv data files

If there is a problem with creating the job, such as incorrect file types, the post request will return an error messages. Otherwise if the job creation is successful a job id will be returned.

1.1.2. Getting a report

To retrieve the report from the Performix Reporting tool a get request is used on the `/report` endpoint. This get request must have the id of the report being retrieved given as a url argument.

If the report job is incomplete the status of the job is returned. If the report job is done then the report document is returned.

1.1.3. Swagger Usage

Accessing `/swagger` endpoint of the Performix API within a browser opens the swagger interface that provides an interactive GUI that can be used to trigger the reporting API. Swagger provides an online form to fill in with the required parameters for posting and getting a report along with an authorize button to set up the authentication bearer token.

On the deployed version of Performix this url can be accessed via `http://performix.biometix.com/api/report/sw`

Submit Date	**8 October 2021**
Submitted by	BBX for Biometix Pty Ltd
Contact	Ms. Somya Singh m: 0412802334 e: s.singh@BBX.com

\

Version History			
Version number	**Description of change**	**Author**	**Date**
V1.0	Initial report release	Somya Singh	30 September 2021
V1.1	Table 10 in section 6.1.3: Typing error has been corrected. The rates in this table did not affect the analysis outcomes	Somya Singh	8th October 2021

1.

2. Terms and Definitions

The terms and definitions identified below are used in this test report.

2. Table 1 Terms and definitions

   Term Abbreviation Definitions

   ---

   attack potential Measure of the capability to attack an IUT (TOE) given the attacker's knowledge, proficiency, resources and motivation attack type Element and characteristic of a presentation attack, including PAI species, concealer or impostor attack, degree of supervision, and method of interaction with the capture device Bona-fide presentations Presentations conforming to ANZ specifications i.e., genuine subject authentication under ideal quality conditions against their own voiceprint template. Concealer attacks When an actor attempts to subvert the system by concealing that their biometric is enrolled in the system Identification attacks When an actor is attempting to be identified in a one-to-many search of a database Enrolment attacks When an actor attempts to enrol a non-live characteristic for the purpose of subverting the system attack presentation classification error rate APCER Proportion of attack presentations using the same PAI species incorrectly classified as bona fide presentations in a specific scenario bona fide presentation classification error rate BPCER Proportion of bona fide presentations incorrectly classified as presentation attacks in a specific scenario presentation attack PA Presentation to the biometric data capture subsystem with the goal of interfering with the operation of the biometric system presentation attack detection PAD Automated determination of a presentation attack presentation attack instrument PAI Object used in a presentation attack Presentation attack species PAS Class of presentation attack instruments created using a common production method and based on different biometric characteristics PAI series Presentation attack instruments based on a common medium and production method and a single biometric characteristic source Implementation under test IUT That which implements the standard(s) being tested Subject The person from whom the biometric enrolment was taken. The target of the attack Target of evaluation TOE Within Common Criteria, the IT product that is the subject of the evaluation. Note: The TOE in Common Criteria evaluations is the equivalent of IUT in biometric evaluations. test approach Totality of considerations and factors involved in PAD evaluation Tester The person performing the simulated PAD attack.

3. Executive Summary

BBX undertook the testing of xxx compliance with the specifications of ISO/IEC 30107-1 and ISO/IEC 30107-3. The evaluation was conducted between xxx

As established in the test plan shared previously, evaluation was conducted in a technology evaluation manner where a test corpus was prepared with simulated enrolments, level A playback attacks and bona-fide presentations, and samples in enrolment dataset were compared against the samples in the verification dataset.

The level − 1 presentation attack detection evaluation was undertaken in accordance with the customer requirements for [Solution] [Vendor] NSS version 12 verification sub-component only. The presentation attacks designed to conduct this evaluation required basic to enhanced basic expertise of the attacker, enhanced basic knowledge of the target of evaluation, and window of opportunity, and moderate equipment costs associated with the artefact creation and usage process.

3.1. Purpose, Scope, and Users

This test report provides the findings and recommendations associated with the Level-1 presentation attack detection testing conducted on the [Customer][Solution]– [Vendor] NSS Version 12 solution.

The intended users of this report are the key stakeholders from [Customer]CAT Team.

3.2. Background

ntroduces new capabilities and services to the biometric market, providing a formal and standardised setting for eIDV (electronic identity verification) biometric and identity software application testing, as well as certification services that are compliant with laboratory and biometric testing ISO/IEC standards and NIST accreditation standards.

3.3. Test Constraints

- This evaluation was conducted in accordance with the recommendations of ISO/IEC 30107-3, which does not provide a specific pass or fail criteria or target APCER or APCER. The test results presented in this report are a result of testing conduced in compliance with the respective standards and frameworks within the scope of BBX's NIST/NVLAP accreditation.

- It is essential to note that the unique context of the given PAD mechanism, the collection of PAI species, the application, the test technique, and the tester all influence error rates. Error rates for PAD processes are not always comparable across similar tests, and they are not always repeatable by multiple test facilities.

- BBX evaluated what is considered to be a representative sample of the commercially available system and employed the appropriate test methodology to ensure normative implications of the ISO/IEC 30107-3 standard.

- Only the PAD verification subsystem (playback detect) of the background model version 12 of [Customer][Solution] solution was tested in this evaluation. Hence the report metrics correspond to the ISO 30107 recommended PAD classification systems.

- It is essential to note that the results presented in this report serve as the certification that [Customer][Solution] solution [Vendor] NSS Version 12 has undertaken testing in conformance with the specifications of ISO/IEC 30107-3 standard. As the standard does not have specific pass or fail levels associated with the metrics, this report does not provide a pass or a fail against any target criteria.

3.4. Scope

3.4.1. In Scope

BBX has evaluated the item under test (IUT) in a level 1 presentation attack detection evaluation in compliance with the recommendations of ISO/IEC 30107-3.

Presentation attack detection assessments go into one of three categories: concealer, verification, or identification PAD evaluations. This assessment only looked at verification or authentication presentation threats.

The evaluation was carried out in a technology evaluation format, with the algorithm provided with a test corpus including simulated attacks and actual audio samples of suitable quality. The attack types have been chosen to be appropriate for the target application, as stated in clause 11 of ISO/IEC 30107-3.

Only Level A attacks with a presentation attack detection rating of level -1 were conducted. These attacks were carried out with the subjects' assistance, using commonly accessible materials, and were developed and tested in less than eight hours per subject.

The following sections detail aspects that fall within the scope of BBX testing.

- To test and validate the performance of the [Customer][Solution] - [Vendor] NSS Version 12 solution in a level-1 presentation attack detection testing effort.

- To report of the solution's performance in line with the metrics recommended by ISO/IEC 30107-3.

3.4.2. Out of Scope

- Although attacks on a biometric system can take a variety of forms, such as overriding or modifying the database, changing the reference, and so on, and can be initiated by any actor, ISO 30107 focuses on biometric attacks on the data capture subsystem (e.g., attacks involving a mobile phone camera) by biometric capture subjects attempting to subvert the system's intended operation. The protection of the data capture subsystem, including the sensor, from modification, replacement, or removal, as well as the protection of communication between the data capture subsystem and other subsystems, falls beyond the scope of this evaluation.

- This evaluation does not cover concealer attacks, identification attacks, or enrolment attacks.

- Evaluation of the end-to-end system performance by conducting a scenario evaluation with real human test subjects falls out of the scope of this evaluation.

- Evaluation of the item under test's operational performance in a deployed implementation is out of the scope of this evaluation.

- PAD Level 1 evaluation of [Customer][Solution] [Vendor] NSS Version 12 solution was conducted with simple attack types, where attacker would have an easy access to the target biometric characteristics and require minimal expertise for the creation of the attack. Sophisticated level 2 presentation attacks such as deep fakes, voice mimics, and synthetic speech cases fall out of the scope of this evaluation.

4. Testing Background

The testing methodology provides the background to the target of evaluation and the approach taken for testing. This identifies the incoming requirements to achieve the testing objectives and the test outputs based on limitations identified.

The goal of this evaluation was to measure the presentation attack detection performance of the item under test in alignment with BBX's NIST/NVLAP validated Level 1 methodology. For this primary reason, this evaluation imitated the functional (input to output) and procedural facets of the proposed concept of operation. This was accomplished by –

- Assessing the item under test using 7 Level A presentation attack species and 10 conformant test subjects.
- Employing Level-A face presentation attack instruments in alignment with the target of evaluation – [Vendor] NSS Version 12 sub-component.

4.0.1. Data Privacy and Management

The ISO/IEC 30107-3 complaint PAD level 1 evaluation process requires the exposure of a limited amount of personally identifiable information. To the maximum extent possible BBX seeks to protect privacy of the subjects in our PAD test corpus by:

- Ensuring the minimum required personal data is collected
- Ensuring explicit consent and understanding of the personnel data collected
- Ensuring that all test transactions are performed in a secure test environment
- Recording only the transaction results relevant to the evaluation
- Ensuring that all transaction data is deleted/destroyed/de-identified soon as possible after the completion of the testing process

4.0.2. Internal Documentation

Following BBX documents were utilised in this evaluation:
Version Number Title Abbreviation Date Author

Contractual
PAD Level 1 and Version 12 Analysis Services SoW 18 June 2021 Biometix for BBX Pty Ltd **Procedures**
V1.0 PAD Evaluation Test Method BXL_10_TMP_10.4 BBX Pty Ltd V1.0 Sampling procedure BXL_12_SPP BBX Pty Ltd V1.0 Testing Report Procedure BXL_12_TRP BBX Pty Ltd V1.0 Quality Assurance Procedure BXL_11_QAP BBX Pty Ltd **Policies**
V1.0 Quality Policy BXL_02_QPQ_02.1 BBX Pty Ltd

4.0.3. External Documentation

Following External documents were utilised to inform this evaluation:
Version Number Title Abbreviation Date Author

2020 NIST General Requirements NIST Handbook 150:2020 2020 NIST/NVLAP 2009 NIST Biometric System Testing NIST Handbook 150-25 2009 NIST/NVLAP 2016 Information Technology – Biometric Presentation Attack Detection- Part 1: Framework ISO/IEC 30107-1 2016 ISO/IEC 2017 Information Technology- Biometric Presentation Attack Detection – Part 3: Testing and Reporting ISO/IEC 30107-3 2017 ISO/IEC 2021 Information Technology – Vocabulary – Part 37: Biometrics ISO/IEC 2383-37 2021 ISO/IEC

5. Scenario & Item Under Test

The Item Under Test (IUT) is the [Vendor] NSS Version 12 system sub-component of the ANZ Australia's [Solution] version 12 solution. The table below lists the hardware and software utilised for the data analysis and testing phases:
System Specifications (ANZ Australia)

System Version [Vendor] NSS Version 12
Test Environment [Customer]Test Environment
Application Software (BBX)

Performix Provider Biometix Pty Ltd Version 6.0.1 Jupyter Notebook Provider JupyterLab Version 6.1.4 Microsoft Office Provider Microsoft Pty Ltd Version 2107

It is the responsibility of the owner of the item under test to ensure that the settings align with the objective of the evaluation and no configurations are changed during the evaluation.

BBX ascertained that the software versions are installed and configured appropriately and verification that the system is operating correction before initiation of the testing process was conducted ensured by the concerned laboratory personnel. Network patch status reports were observed monthly during the testing process to ensure no anomalies associated with the BBX internal environment were found.

[Vendor] NSS Version 12 The playback-detect system subcomponent (Item under test) has two facets:

- **Footprint Playback:** Footprint Playback detection identifies if audio buffers belong to the same utterance. The system compares the present audio to previously stored ”footprints” of 5 verification passphrases that were previously collected. The current footprint is marked as a recording if the two footprints are highly similar.

- **Channel Playback:** The existence of signal artefacts generated by the recording and playback processes is detected by channel playback detection, which isolates playback attacks based on an ANZ specified false alarm rate of 2%.

The business rules and order of sample processing means that the footprint scores are not recorded for all playback attempts if channel playback is flagged first.

6. Test Methodology

This evaluation was conducted using the test corpus prepared by BBX personnel. A pseudonymised version of the test corpus was provided to [Customer]to be run in a test environment. Clear instructions regarding data handling were provided to the respective ANZ personnel to ensure untainted corpus and valid test results.

The test corpus prepared by BBX was provided to ANZ to run the test. Instructions on how to run the test and the steps to be followed were provided by BBX to the concerned personnel who ran the test. The test resulted in 50 bona-fide attempts and 350 playback attempts with 7 presentation attack species, based on samples from 10 test subjects.

It is essential to note that the rate of false positives have been measured based on the default decision policy, the playback threshold, and any thresholds for sample quality. The rate of false negatives for each presentation attack species has been reported with these details, alongside the estimated bona-fide presentation classification error rate at the same values.

6.1. Pre-Testing

This section provides a description of the pre-testing activities that were conducted to inform the final test run. BBX confirmed our understanding of how the [Vendor] NSS Version 12works and our initial methodology to test it as mentioned below:

- BBX generates and provides audio samples in .wav formats with appropriate file names as labels including both enrolments (x3) for the de-identified test subjects and a number of verification attempts (including bona fide and presentation attacks)

- ANZ would enrol the speakers using the three enrolment audio files – letting BBX know if any of them have issues so we can reissue as clean enrolments for each speaker are expected for this evaluation

- ANZ would generate and provide results by matching the verification attempt audio files to the previously enrolled speakers

- The output data would include on a minimum Playback Footprint (PBF) score, Playback Channel (PBC) scores, and Biometric Scores associated with each transaction

- A dry run was conducted following the understanding of the [Vendor] NSS Version 12and testing methodology by providing ANZ with audio files corresponding to two unique test subjects.

Based on the initial testing methodology described above, the following pre-testing activities took place:

- The first sample corpus was provided to ANZ for a dry run-on the 30th of June and further correspondence was received regarding the required format of the Test Corpus and anonymised metadata csv on the 7th of July.
- BBX re-formatted the test corpus in the structure ingestible by the ANZ test system and made these available on the 13th of July.
- Advise was received from ANZ on the 19th of July to include ".wav" extension in the encrypted filenames provided, provision of mono, not stereo files (as stereo files are not accepted by NSS), and "True/False" for enrolment file column. This was actioned by BBX on the 19th of July
- First sample run output was received on the 19th of July and analysed by BBX as part of the pre-testing process. It was noted that the Footprint Playback Scores (PBF) were not generated for the runs conducted. Upon further understanding of the order of sequence of different subcomponents including footprint playback detect, channel playback detect, and other final decision reasons, it was noted that Footprint playback is only performed when the voiceprints have received a match and passed channel playback detection.
- As [Vendor] NSS Version 12system comprises of both footprint and channel [Vendor] NSS Version 12 it was necessary to receive footprint playback scores for this evaluation. Hence, the following approach was utilised to get the footprint playback scores:
- **Revised Footprint Score Generation approach:**
- To receive the footprint playback, detect scores, all other checks had to be disabled and a target false accept rate of 100% was set so PBF scores could be triggered.
- The analysis approach was revised to only consider system thresholds corresponding to biometric score, playback, etc. so the transactions that would have received a match, non-match, inconclusive, quality flags etc., if the system were set up normally could be traced during the analysis process.
- Provision of 3 enrolments and 5 verifications (bona-fide/genuine attempts) to perform bona-fide testing baseline and 'train' the footprint playback was recommended followed by the provision of all playback attack audio files. This was because footprint playback compares the audios received from the previous 5 transactions to generate a PBF score. BBX understanding of the processing order of footprint [Vendor] NSS Version 12is provided below:

4. Figure 1 Order of processing for Footprint playback score generation

Using the footprint [Vendor] NSS Version 12order of processing, the dataset was run in the manner that for each presentation attack conducted only the previous 5 bona-fide audio samples were used for utterance comparison. Figure shown below provides the order of sequence of playback attacks for footprint detect:

Based on the outcomes of the pre-testing process, the final test corpus was made available to ANZ on the 2nd of August and resulting output logs were received on the 13th of August for analysis.

6.2. Order of Use

As transactions in the deployed end-to-end system are executed separately, the methodology for the execution of this evaluation required the completion of enrolment before commencing the verification. This is because the authenticating customers separately use the end-to-end system (of which [Vendor] NSS Version 12is a subpart) and perform a transaction in a sequential manner (enrolment followed by authentication).

Hence, the ANZ personnel running the tests were advised to conduct the tests in a sequential manner where they first enrol the voiceprints provided followed by the verification run.

It is understood that mated-comparisons of genuine enrolment templates against bona-fide, and playback attack presentations, in the test corpus for each (ordered) pair is likely to exhibit low degree of correlation in the transactions (nearly independent comparisons). Hence, the same test run approach was advised to be followed for both bona-fide and playback voiceprints (i.e., both bona-fide and playback voiceprints were provided together with correctly labelled files for traceability).

Upon further understanding of the order of sequence of different subcomponents including footprint playback detect, channel playback detect, and other final decision reasons, it was noted that footprint playback is only performed when a voiceprint receives a match and has passed channel playback detections. Based on this, a revised testing approach was followed for receiving [Vendor] NSS Version 12scores associated with each comparison as described in the section above.

6.3. Test Environment Set-up

This section provides a summary of test environment set-up during the data collection process to ensure the datasets prepared for this evaluation were representative of the target population and were collected in close to production like environment.

6.3.1. Datasets

- The datasets prepared for this evaluation were made up of audio samples from 10 conformant and demographically diverse test subjects. BBX ensured written consent was acquired from the participating candidate before the commencement of the data collection process.
- The audio samples varied in the sense that the subjects were allowed to hold the recording device with one hand or two hands, speaking too close to the phone or too far etc.
- All audio samples were collected in a controlled laboratory environment using a single acquisition device (mobile phone – iPhone version 14.5). The subjects were provided with guidance and instructions by the BBX data acquisition technician on how to capture their audio sample.

6.4. Test Corpus

This section aims to specify the high-level attributes of the test corpus based on the requirements established. These attributes include but are not limited to age, gender, ethnicity, and other factors under which the corpus was collected.

A small sample test corpus was prepared and provided to [Customer]containing audio files from two test subjects to inform the dataset formats and ensure traceability of the results received in the final test run.

Post the finalisation of the testing methodology to be utilised to run the tests and for the data analysis, test corpus containing the following samples was made available to ANZ for a final run:

- **Enrolment:** Three enrolment audio samples were provided per person. This was because the current decision policy of the [Solution] solution allows the enrolee to provide three audio samples to successful enrol their voiceprint in the solution.
- **Verification:** Two categories of verification samples for each subject were made available:
  - **Bona-fide:** 5 valid voiceprints per person in alignment with the system's verification decision policy were provided. Each one of the 5 genuine audio samples represented one attempt.
  - **Playback:** 7 different types of playback audios were provided for each user. Each playback type comprised 5 playback attacks per person, corresponding to 5 tries per person per presentation attack instrument type. No playback sample was used more than once.

A summary of the datasets made available to ANZ is provided in the table below:

Table 2 Data fields and Identifiers corresponding to datasets in the test corpus

Subject Enrolment Verification

`**Bona-fide**`

Subject SubjectID_0_E_backgroundNoise_GenuineAuthentic_InstanceID SubjectID_1_B_backgroundNoise_Genu
SubjectID_1_presentationAttackInstrumentID_backgroundNoise_GenuineNonAuthentic_InstanceID

The data identifier definitions are listed below:

- **SubjectID**: Pseudonymised and de-identified test subject identity
- **0:** represents an enrolment audio sample

- **1:** represents a verification audio sample (bonafide presentation or Level A presentation attack)
- **BackgroundNoise:** represents the background noise in the recording environment in decibels
- **GenuineAuthentic:** represents a genuine enrolment and verification transaction
- **GenuinNonAuthentic:** represents a playback attack
- **InstanceID:** attempt number corresponding to a session

A test corpus made up of 10 unique, and conformant test subjects was identified for this evaluation that meets the requirements stated in BBX's NIST validated presentation attack detection test methodology written in compliance with ISO/IEC 30107-1, and ISO/IEC 30107-3. Summary of the test crew demographics in this evaluation is provided in the tables below:

5. Table 3 Age Distributions

   Age Range ISO 19795-5 Recommended Distribution BBX Sourced Test Crew

   < 18 y/o[1] N/A - 18 – 30 y/o 25 – 40% 8 31 – 50 y/o 25 – 40% - 51 – 70 y/o 25 – 40% 2 > 70 y/o N/A[2] -

6. Table 4 Gender Distributions

   Gender ISO 19795-5 Recommended Distribution BBX Sourced Test Crew

   Male 40 – 60% 4 Female 40 – 60% 6

7. Table 5 Ethnicity Distributions

   Ethnicity[3] Biometix Test Plan Proposed Test Crew Inclusion

For this evaluation, the above-mentioned data was recorded for each participating test subject. The anonymised information associated with the metadata recorded remains within the secured BBX information management system and can be destroyed upon request. Any other subject meta-data that would not normally be available to the item under test was excluded from the logs analysed.

6.5. Presentation Attack Instrument (PAI) Species

This Level – 1 presentation attack detection evaluation was conducted using Level A presentation attack instruments.

These presentation attack instruments (PAIs) were selected based on three factors as described in our NIST validated presentation attack detection testing methodology (BXL_TMP_10.4_PAD_Evaluation Test Method):

- **Type:** A designation of the artefact defined by its properties and origin.
- **Access to biometric characteristics:** The relative ease of access to suitable sources from which the artefact can be produced
- **Equipment/Cost:** The relative difficulty and expense to produce the artefact.

These factors enable the tester to gauge the attack potential of the attack instruments used to assess the presentation attack detection performance of the item under test against a certain level of evaluation.

[1] The ISO 19795-5 standard does not include a proportion of test subjects under 18 y/o

[2] The ISO 19795-5 standard does not include a proportion of test subjects over 70 y/o

[3] Adapted from Australian Standard Classification of Cultural and Ethnic Groups (2019) available at https://www.abs.gov.au/ausstats/abs@.nsf/mf/1249.0

11. Table 6 Presentation Attack Species Assessment

    Document Type Score Description

    Type Simple Typically, two-dimensional and/or repurposed from another source

```
**Specialised**     **Typically, three-dimensional, and/or
**Sophisticated**   **Specifically produced, sophisticated
```

    Access to biometric characteristics Easy Publicly or commonly available, usually without knowledge of the target

```
**Moderate**        **Usually requires cooperation of the
**Difficult**       **Usually requires multiple sources wh
```

    Equipment/cost Low Can be produced with standard materials using office or home equipment

```
**Medium**          **May require the use of generic suppl
**High**            **May require the use of specialised s
```

When a presentation attack instrument is classified, it is defined by the highest score achieved in the three categories as either:

- **Level A**: Typically, low effort attackers. More likely to be genuine customers meddling with the system.
- **Level B**: More experienced attackers actively trying to fraud the system.
- **Level C**: Targeted attackers with significant resources, typically with knowledge of biometrics or the system.

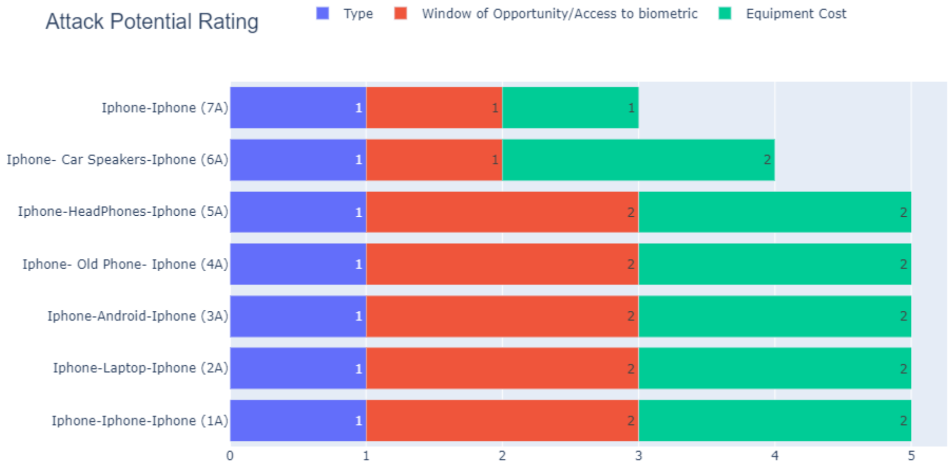
All level A attack instruments were different types of playback attacks created using high quality mono audio samples greater than ~3 seconds under controlled laboratory environment. These instruments are considered appropriate to evaluate the item under test – ANZ Australia's [Solution] [Vendor] NSS Version 12subsystem (version 12) against the ISO/IEC 30107-3 requirements and in a PAD-Level 1 evaluation.

12. Table 7 Level A presentation attack instruments selection criteria

    Score Description

    Type Simple Access to biometric characteristics Easy Equipment/cost Low PAI Classification Level A

Figure below shows the attack potential rating of the attack instrument species utilised in this evaluation:



13. Figure 2 Attack potential rating for the PAI Species utilised in this evaluation

6.5.1. Replay Attack

A replay attack is the use of a recording device to record the common passphrase from a test subject and then to replay that audio against the identity of during verification.

6.5.2. Resources

The data collection process was completed with a recording sample taken in a controlled laboratory environment with the use of a latest iPhone device. The device used for enrolment will be a common iPhone shared amongst the subjects.

- Each test subject recorded 3 enrolment audio samples as per the decision policy of the item under test.
- Each test subject provided 5 bona-fide audio samples – these samples will be collected in the fashion intended by the policy of the item under test.
- Each test subject provided a recording which was used to create the presentation attack instruments corresponding to seven level A presentation attack species.

A replay attack was simulated by playing the recording back with a range of instruments, for instance, a set of headphones used to nest the iPhone during the playback, another iPhone, etc. The sound level of the playback was at normal conversational volume at approximately 50cm from the subject.

All participating test subjects were free from any illness or condition which may significantly alter their normal voice pattern.

14. Figure 3 An overview of the Presentation Attack Instruments development procedure

PAS -1 Playback attacks that fall under presentation attack species 1 (PAS-1) were created by collecting an audio sample from the test subject using an iPhone version 14.5. This audio sample was played back to another iPhone version 14.5, and then replaying the second recording to another iPhone version 14.5

This Level A attack simulated an attacker who recorded a genuine customer's passphrase on an iPhone and is trying to replay the recording to authenticate against them.

PAS-2 Playback attacks corresponding to presentation attack species 1 (PAS-2) were created by collecting an audio sample from the test subject using an iPhone version 14.5. This audio sample was played back to a laptop (Lenovo – Windows 10 OS version
10.0.19043), and then replaying the second recording to another iPhone version 14.5

This attack simulated an attacker who recorded a genuine customer's passphrase on a laptop and is trying to replay the recording to authenticate against them.

PAS-3 Presentation attack species 3 (PAS-3) presentation attacks were created by collecting an audio sample from the test subject using an iPhone version 14.5. This audio sample was played back to an Android version 11, and then replaying the second recording to another iPhone version 14.5.

This attack simulated an attacker who recorded a genuine customer's passphrase on an Android Device and is trying to replay the recording to authenticate against them.

PAS-4 Presentation attack species 3 (PAS-4) presentation attacks were created by collecting an audio sample from the test subject using an iPhone version 14.5. This audio sample was played back to an old iPhone 6 version 12, and then replaying the second recording to another iPhone version 14.5

This attack simulated an attacker who recorded a genuine customer's passphrase on an old iPhone Device and is trying to replay the recording to authenticate against them.

PAS-5 Presentation attack species 3 (PAS-5) presentation attacks were created by collecting an audio sample from the test subject using an iPhone version 14.5. This audio sample was played back to Headphones (Jabra – 34DF2A), and then replaying the second recording to another iPhone version 14.5

This attack simulated an attacker who recorded a genuine customer's passphrase while they were using their headphones and is trying to replay the recording to authenticate against them.

PAS -6 Presentation attack species 3 (PAS-6) presentation attacks were created by collecting an audio sample from the test subject using an iPhone version 14.5. Playing back this audio sample inside a car, and then recording the second audio sample on another iPhone version 14.5

This attack simulated an attacker attempting a replay attack in a car.

PAS-7 Presentation attack species 3 (PAS-7) presentation attacks were created by collecting an audio sample from the test subject using an iPhone version 14.5, and playing back this recording on another iPhone version 14.5

This presentation attack species represents firsthand replay attack.

7. Performance Analysis

BBX performed the evaluation of the [Solution] version 12 [Vendor] NSS Version 12solution using 7 level A presentation attack species described in section 5.5 of this report. This section provides a breakdown of successful and unsuccessful test cases based on the attack potential associated with each type of attack vector covered in the testing. The figure below shows the distribution of biometric scores for attack presentations against scores for bona-fide presentations. It was observed that genuine authentic transactions had notably higher biometric scores, which is consistent with an understanding of the improved quality of natural speech against a recording.

Figure 4 Distribution of biometric scores for Attack presentations (TEST_NAME: GNA) Vs. Bona-Fide presentations (TEST_NAME: GA)

7.1. Attack Presentation

The presentation attacks for all test crew participants corresponded to 5 presentations per species. This resulted in 50 presentation attack attempts against each species, with some processing errors on speaker 2 and 7 making up a total of 337 valid data points for analysis

According to ISO/IEC 30107-3, Imposter Attack Presentation Match Rate (IAPMR) is defined as the proportion of imposter attack presentations using the same PAI species in which the target reference is matched. The standard defines Attack Presentation Classification Error Rate (APCER) as the proportion of attack presentations using the same PAI species incorrectly classified as bona-fide presentations in a specific scenario. The table shown below lists the individual IAPMRs and APCERs against each PAI species utilised.

Note that the rates mentioned below exclude the attack attempts receiving an audio too short flag for a close to real-world performance overview of the solution (including the attempts flagged as audio too short will correspond to marginally lower IAPMRs against each attack type). These are deemed as quality errors and identified in the Attack Presentation Non-Response Rate (APNRR).

While insufficient biometric quality may cause an attack instrument to fail in PAD testing, there is no reason to anticipate a specific level of quality from artefacts in general. Artefact samples have the potential to be of greater quality than human biometric samples. In the absence of an attacker skill model, it is accepted to assume a ”worst case” scenario in which the attacker always utilises the highest possible quality. This way, at the very least, a guaranteed minimal detection rate for the given test set may be determined.

15. Table 8 Summary of Presentation Attack Detection performance of [Solution] Version 12 playback detect

Attack Species (Playback Attacks)	**Presentation Attacks**[4]	**IAPMR**[5]	**APCER**[6]
iPhone-iPhone-iPhone	**45**	0%	0%
iPhone-Laptop-iPhone	**30**	0%	0%
iPhone-Android-iPhone	**29**	0%	0%
iPhone-Old Phone-iPhone	**37**	0%	0%
iPhone-Headphones-iPhone	**20**	0%	0%
iPhone-Car Speakers-iPhone	**47**	0%	0%

[4] Excluding attack attempts flagged as audio too short

[5] Includes proportion of arrack attempts receiving a match/ (total attack attempts-quality errors)

[6] Includes proportion of attack attempts receiving match, mismatch, inconclusive/ (total attack attempts-quality errors)

iPhone-iPhone 49 0% 0%
Overall 257 0% 0%

```
**APCER Chann
0.39%
```

Where an overall APCER of 0% was noted for all PAI species, the APCERs' associated with the individual performance of footprint [Vendor] NSS Version 12and channel [Vendor] NSS Version 12were greater than zero. The two subsystems together however produced an APCER, and therefore IAPMR, of zero.

Channel [Vendor] NSS Version 12system sub-component was found to be far more effective in identifying playback attacks than footprint detect. This is the optimal outcome as channel playback also represents a lower proportion of bona fide rejections and is also more effective in a wider range of use cases where footprints are not useful, such as on the enrolment process

Attack presentation classification error rates greater than 0% for a PAI species only demonstrate that the PAI is capable of resulting in a successful attack. A different tester's attack presentation classification error rate might be greater or lower. Furthermore, training to recognise key material and presentation characteristics may enhance the attack presentation classification error rate for the same PAI species. The tester's expertise and knowledge, as well as the availability of the requisite resources, are important elements in PAD testing and are considered when performing comparisons or performance analysis (see Figure 2).

7.1.1. Attack presentation non-response rate

This section describes the performance of the solution in relation to the proportion of presentation attacks using the same PAI species that result in no response at the PAD subsystem. In this evaluation, a number of "*Audio Too Short*" flags were noted for attack presentations corresponding to no response from playback detect.

Table below provides the APNRRs associated with each PAI species utilised in this evaluation.

Table 9 Attack Presentation Non-Response Rates corresponding to PAI species

Attack Species (Playback Attacks) Presentation Attacks APNRR

iPhone-iPhone-iPhone 49 8.16% **iPhone-iPhone** 50 2%

PAI species 5A (iPhone-Headphones-iPhone) had the highest non-response rate subject to Audio Too Short quality error with the lowest non-response rate for iPhone-iPhone playback attack species (7A).

7.1.2. Attack Presentation Acquisition Rate

Attack Presentation Acquisition Rate (APAR) is described as the proportion of attack presentations using the same PAI species from which the data capture subsystem acquires a biometric sample of sufficient quality.

The APARs associated with each presentation attack instrument type are provided below:

Attack Species (Playback Attacks) Presentation Attacks APAR

iPhone-iPhone-iPhone 50 98% **iPhone-Laptop-iPhone** 50 98%

7.1.3. Correct Attack Classification

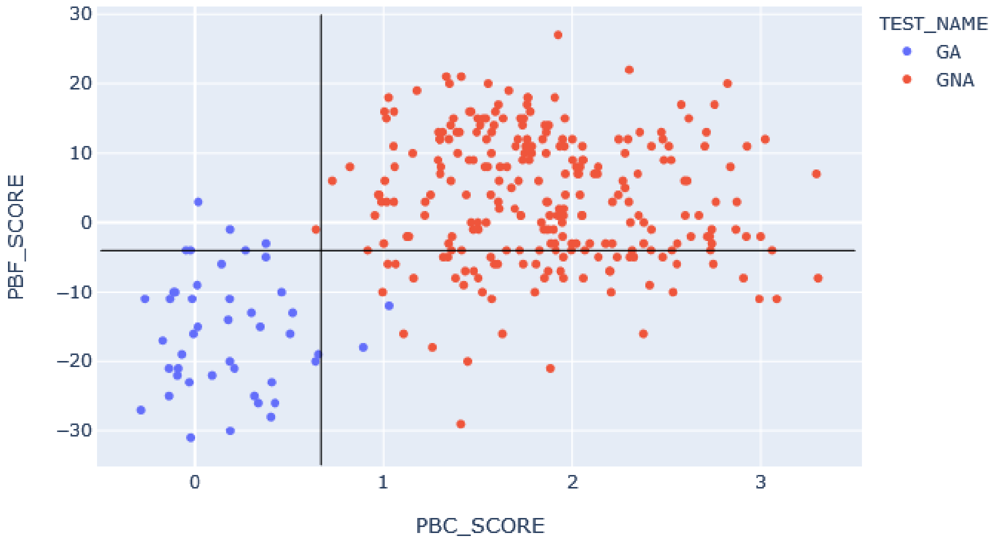
The correct attack classification rate for footprint and channel [Vendor] NSS Version 12was calculated based on the current system thresholds for these respectively. For a closer to real-world understanding of the system performance attempts with net audio of less than 1 second were excluded from the analysis. This resulted in a total of 459 presentation attack attempts. Note that the footprint and channel scores were generated separately with different testing approach utilised for both as described in section 5.1 of this report. Hence the table below lists the individual performance of each against the total number of valid presentation attacks performed and there is no correlation between footprint and channel detect for the sets analysed.

19. Table 10 Individual performance of channel and footprint [Vendor] NSS Version 12against valid playback attack presentations

Playback Detect Percentage of attempts correctly classified as playbacks[7] Threshold

Footprint 209/257 81.3% -4[8] Channel 256/257 99.6% 0.67

Distribution of channel playback scores and footprint playback scores associated with attack presentations is provided in the figure below, with the bona fide also indicated in blue for comparison and the respective thresholds indicated by the black line. Note the observed difference in performance and the strong separation between bona fide and presentation attacks.



22. Figure 5 PBF Scores Vs. PBC Scores

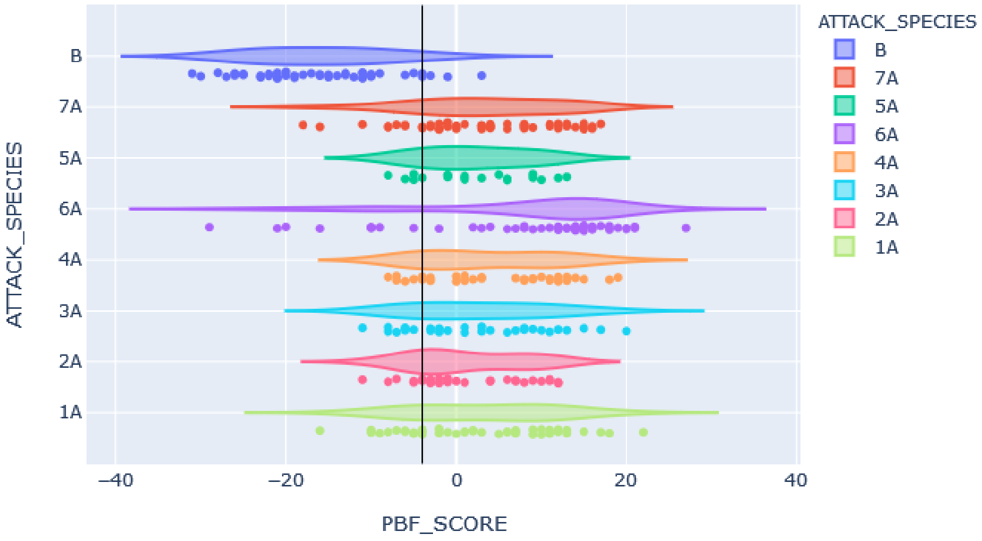
Where the performance of channel playback was highly effective in detecting attack presentations, a high proportion of attack presentations received lower than threshold footprint playback scores.

7.1.4. Footprint Playback Detect

Figure below shows the distribution of footprint playback scores against each playback presentation attack instrument type utilised in this evaluation. For a close to real-world performance analysis, it was assumed that an attacker would make all possible efforts to ensure that the quality of the attack is optimal, hence, attempts flagged as "*Audio Too Short*" were excluded. Playback attacks where user's voice was recorded on an iPhone and played back through car speakers into another iPhone (6A), and attacks using iPhone-iPhone-iPhone combination (1A) where the most successful attack types for footprint playback detect. Footprint [Vendor] NSS Version 12had the best performance against iPhone-Headphone-iPhone (5A) playback attack types. Bona Fide is indicated in dark blue by 'B'.

[7] i.e., Above the playback threshold.

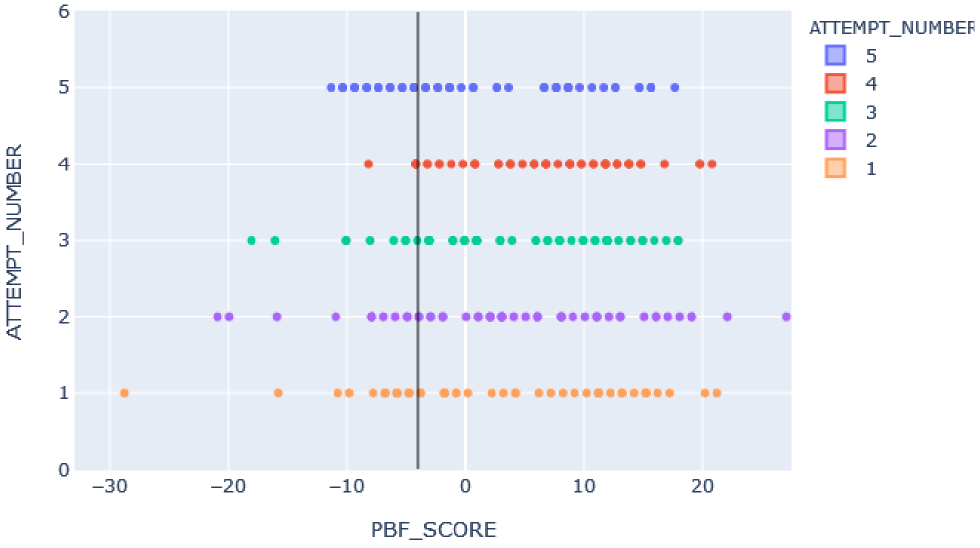
[8] Current version 12 PBF threshold @ biometric threshold -1.01 and FAR 0.05%



23. Figure 6 Distribution of Footprint playback scores for valid attack presentations categorised by attack species

Footprint playback detection identifies if audio buffers belong to the same utterance by comparing the present audio to 5 previously stored footprints of the common passphrase from the same user and, marking them as a "recording/playback" if the footprints have a high degree of similarity, i.e., lower the similarity, higher the probability of an audio not receiving a footprint playback flag.

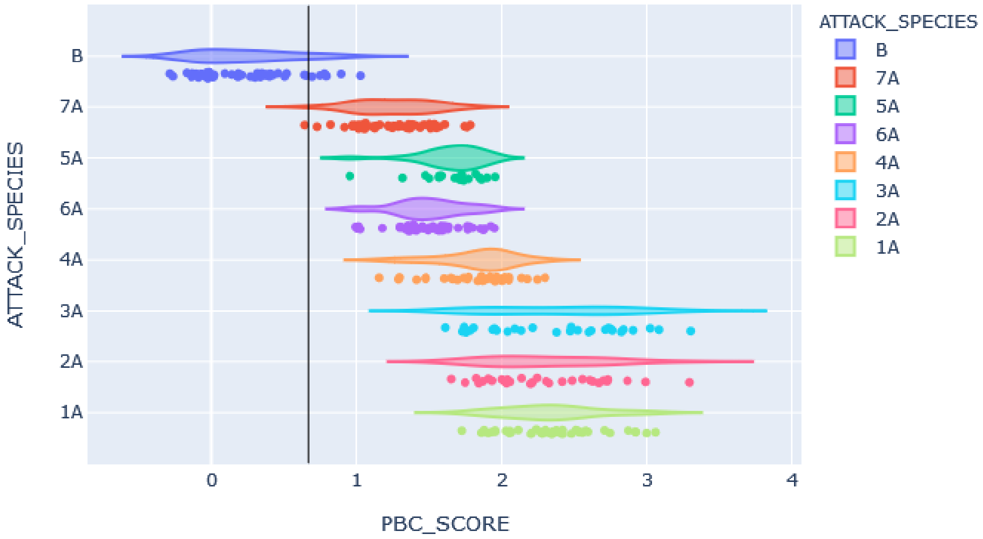
Figure below shows the effect of attempt number on footprint playback scores for presentation attacks conducted. Performance of footprint detect was noted to improve marginally on the 5th attempt vs the 1st attempt of the same presentation attack type. This is consistent with the footprints of previous attacks informing the decision made.



24. Figure 7 Effect of attempt number on the distribution of Footprint Playback scores for valid playback attacks

7.1.5. Channel Playback Detect

The distribution of channel playback scores against each playback presentation attack instrument type is shown in the figure below with the bona fide also indicated in dark blue as 'B'. Performance was substantially improved over footprint playback detection with fewer rejects, more closely grouped performance, and a greater separation between distributions of bona fides against presentation attacks.



25. Figure 8 Distribution of Channel playback scores for attack presentations categorised by attack species

7.2. Bona-fide Presentation

This section describes the performance of the [Solution] version 12 [Vendor] NSS Version 12system in relation to its performance against Bona-fide presentations. For this evaluation, Bona-fide presentations were those that conform to ANZ Specifications with optimal audio length, genuine user samples provided under close to real world conditions e.g., average SNR associated with bona-fide samples was 30.

For this evaluation, 50 Bona-fide attempts were conducted using 5 bona-fide samples from each unique user with 10 test subjects.

7.2.1. Bona-fide Presentation Classification Error Rate (BPCER)

ISO 30107 describes the Bona-fide Presentation Classification Error Rate (BPCER) as the proportion of bona-fide presentations incorrectly classified as presentation attacks in a specific scenario. Table below, provides a summary of BPCERs associated with footprint and channel playback detect. It was noted that the false playback indications were higher than what has been observed in other testing. For the purposes of this evaluation, the presence of increased classification errors serves to provide a set of speakers that would benefit more greatly from a reduction in the playback thresholds

Table 11 Footprint and channel BPCERs

Playback Detect Bona-fide attempts BPCER Threshold

Footprint 50 22% -4[9] **Channel 50 14% 0.67 Combined 50 26% N/A**

7.2.2. Bona-fide Presentation Non-response Rate (BPNRR)

No non-responses were observed for bona-presentations owing to the quality errors.

7.2.3. Failures to Acquire & Failure to Enrol

Where no Failures to Acquire were measured for this evaluation, it was noted that the audios for Test subjects 2 and 7 did not enrol successfully.
Speaker 2 audio is OK, PLAYBACK_INDICATION, PLAYBACK_INDICATION.

[9] Current version 12 PBF threshold @ biometric threshold -1.01 and FAR 0.05%

Speaker 7 audio is OK, OK, PLAYBACK_INDICATION.
As a result, 13 presentation attack attempts were not recorded of the 350 attempted.

7.2.4. Modelled Performance

Prediction of the generalised performance of the current version 10 BGM against modelled PBF thresholds was made using bona-fide utterance 1 transactions from the PAD testing conducted. The examples of possible configuration options are listed in the tables below for Footprint and Channel [Vendor] NSS Version 12system sub-components individually.

Playback thresholds were adjusted to provide an APCER of 0.39%, corresponding to one successful presentation attack in the course of this evaluation. This was determined as the single passing attempt exceeded the match score threshold significantly and sat close to the PBF threshold. By adjusting both the footprint and channel threshold together, the maximum performance can be gained. This was because there was minimal cross over between the classification errors between the two technologies

27. Table 12 Modelled generalised verification rates against potential changes to footprint playback threshold - version 12 BGM (these are session rates)

    Footprint Playback Threshold Generalised Verification Rate

    ---

    -4[10] **- Current Setting 74%** -3 – Reduced Setting 80% -2 – Reduced Setting (Rate adjusted for this evaluation) 82%

28. Table 13 Modelled generalised verification rates against potential changes to channel playback threshold - version 12 BGM

    Channel Playback Threshold Generalised Verification Rate

    0.67 – Current Setting (2% False Alarm Rate) 74% 0.80 (1% False Alarm Rate) 74% 0.99 (Rate adjusted for this evaluation) 76%

29. Table 14: Combined Rates

    ```
                    Generalised Verification Rate   Attack Presentation Classification Error Rate
    ```

    ```
                    **Footprint -4**                **Footprint -3**
    ```

    Channel – 0.67 74% 0% 80% 0% 82% 0% **Channel – 0.80** 74% 0% 80% 0% 82% 0% **Channel - 0.99** 76% 0% 82% 0% 84% 0.39%

The figure below provides a comparison of the bona fide and attack presentations, measured against the current and potential threshold (-4 PBF, 0.99 PBC).

31. Figure 9: Comparison of bona fide and attack presentations with thresholds indicated

8. Deviations and Exclusions

ISO/IEC 30107-3 covers a number of attack types, system operational types, and evaluation techniques. This report certifies only the following items tested:

- [Customer][Solution] Version 12 [Vendor] NSS Version 12algorithm performance
- Attacks involving simple playback attacks
- Evaluation of the PAD Classification subsystem

BBX has undertaken every step to ensure no deviations or omissions from the ISO/IEC 30107 standard were made.

[10] Current version 10 PBF threshold @ biometric threshold 2.68 and FAR 0.05%

9. Findings and Recommendations

Recommendations

BBX has completed Level 1 PAD testing for the [Customer][Solution] version 12 [Vendor] NSS Version 12system. The purpose of this report is to report on the testing that was done as well as the metrics that were gathered as an outcome of the testing. This evaluation did not assess adherence to any target criteria.

The overall system meets all of the normative requirements for Level 1 testing with ISO/IEC 30107-3, according to section 6.0 ”Performance Analysis.”

BBX confirms that [Customer]Version 12 [Vendor] NSS Version 12system meets the level 1 criteria for presentation attack detection.

Constraints

As described in section 2.3 of this report, BBX has evaluation what it believes to be a representative sample of the commercially available solution with the utilisation of appropriate testing methodology stemming from the specifications of ISO/IEC 30107-3 standard.

The results associated with the PAD testing of the [Customer][Solution] version 12 [Vendor] NSS Version 12solution have been reported in section 6.0 according to the testing methodology presented in section 5.0 of this report.

It is important to note that as per section 2.3, the results presented in this report serve as certification that [Customer][Solution] solution version 12 [Vendor] NSS Version 12has undergone testing in accordance with the ISO/IEC 30107-3 standard standards. Because the standard does not provide pass or fail levels for the metrics, this report does not indicate a pass or fail against any target criteria.

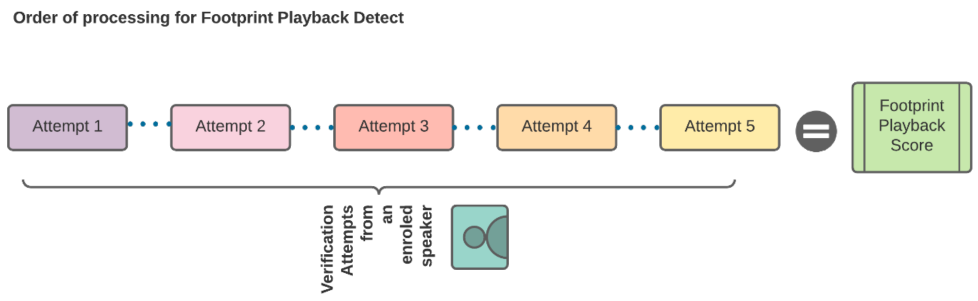
Deviations

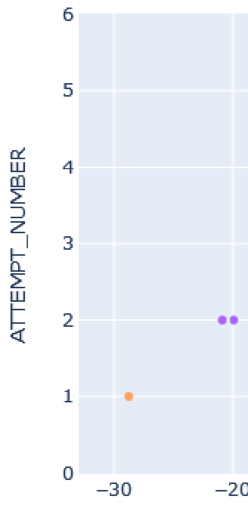
BBX has taken every measure to ensure that no deviations were made from the specifications of the standard.

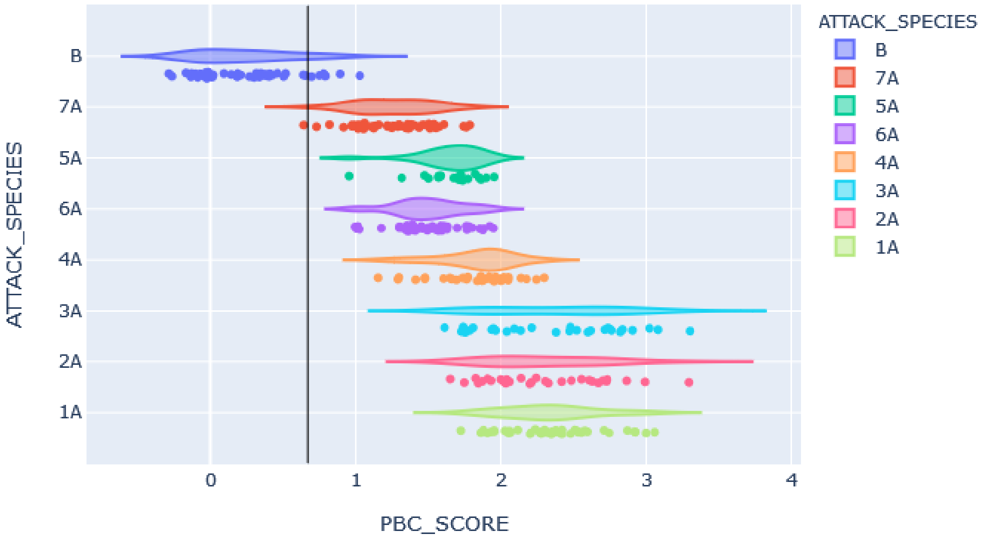
Conclusions

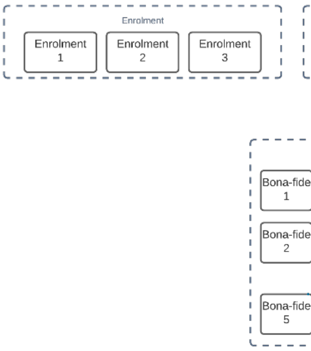
There are no other comments or thoughts from BBX that are not addressed in this report.

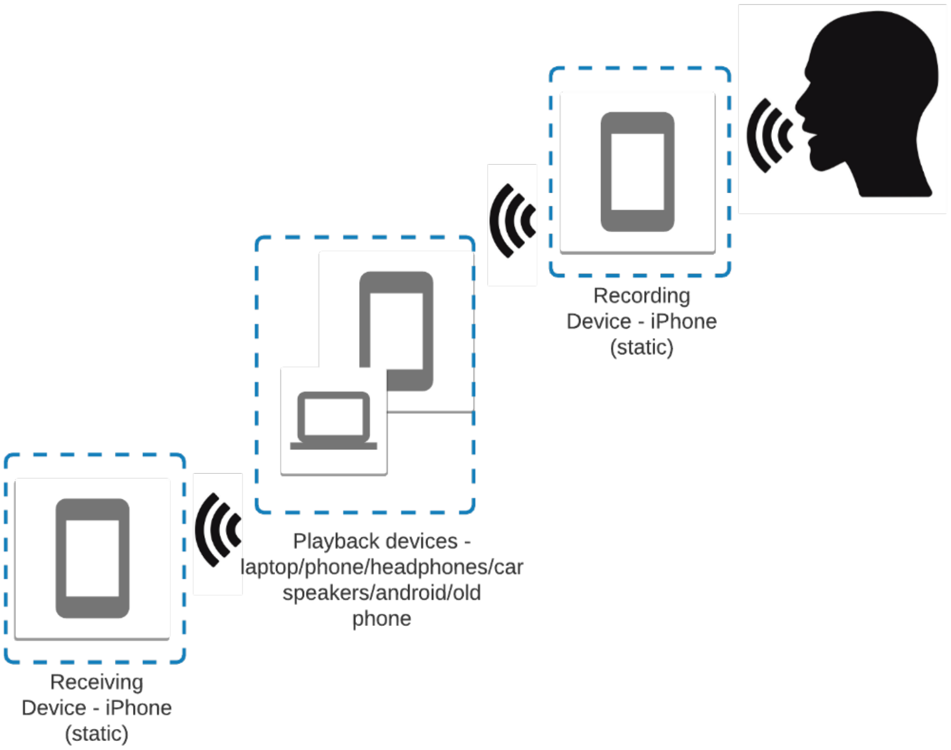
Dr. Ted Dunstone
Senior Responsible Officer
ted@BBX.com
30th September 2021

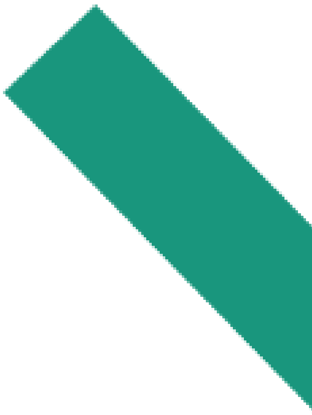

Order of processing for Footprint Playback Detect
Attempt 1
Attempt 2
Attempt 3
Attempt 4
Attempt 5
Footprint Playback Score
Verification Attempts from an enroled speaker

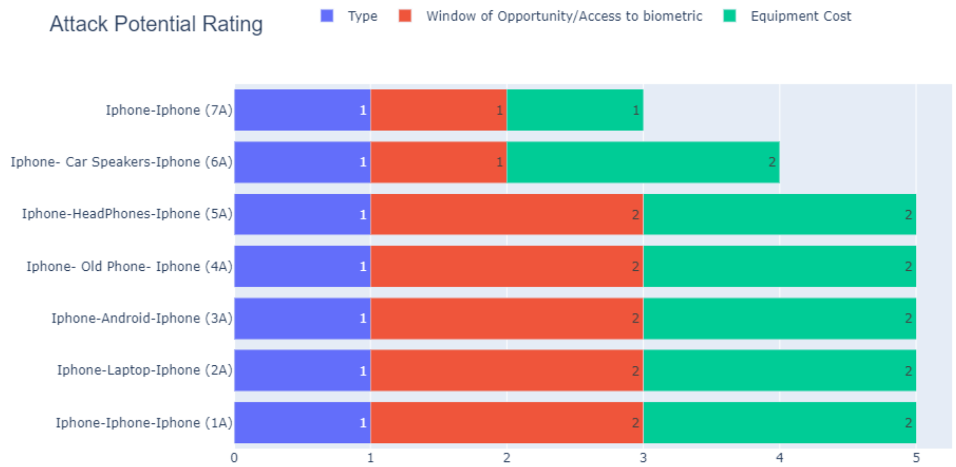


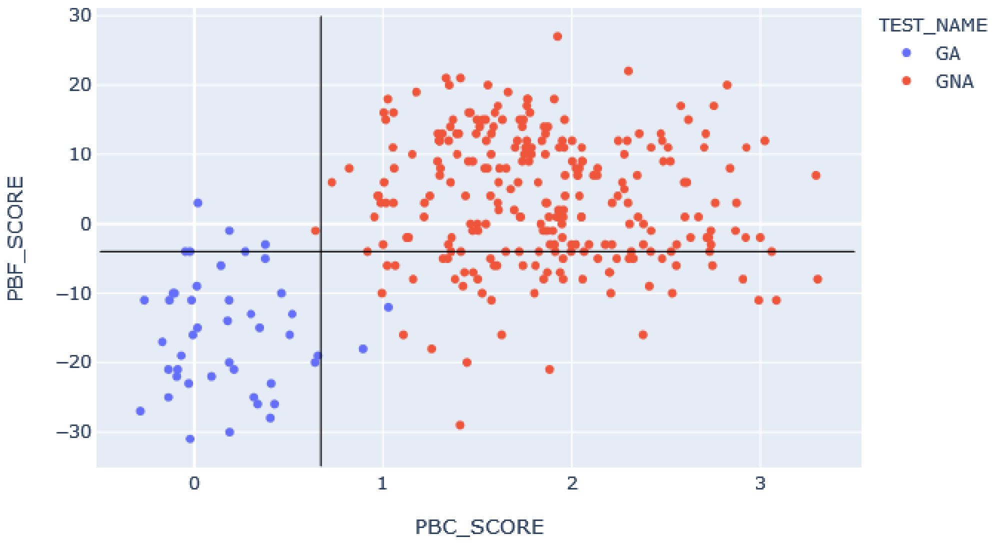

ATTACK_SPECIES
B
7A
5A
6A
4A
3A
2A
1A
0
1
2
3
4
PBC_SCORE
ATTACK_SPECIES
B
7A
5A
6A
4A
3A
2A
1A

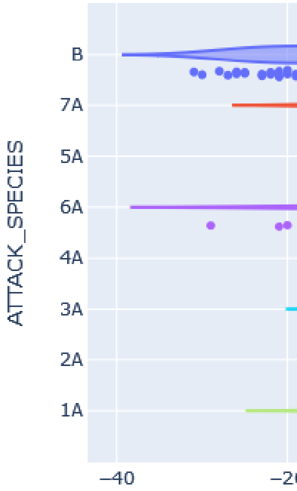

Enrolment
Enrolment 1
Enrolment 2
Enrolment 3
Bona-fide 1
Bona-fide 2
Bona-fide 5


Recording Device - iPhone (static)
Playback devices - laptop/phone/headphones/car speakers/android/old phone
Receiving Device - iPhone (static)









10. Introduction

The increasing role of biometric technologies has led to calls for greater regulation of biometrics and identity. Other countries are also considering how best to regulate these technologies and some have enacted specific regulatory frameworks for biometrics. Given this there is an urgent need to better set out an approach to regulation of biometrics and its expectations of agencies using or proposing to use biometrics and to contribute to the wider discussion about whether existing regulatory frameworks adequately address the risks and maintain the benefits of using biometric technologies.

We argue that it is important to set out the key principles that should underpin the development of a robust regulatory framework for biometrics and to establish the level of oversight that should be provided.

This paper sets out some of the key principles, and makes three recommendations:

– An identity commissioner should be set up with an appropriate remit and powers.

– The commissioner should review the current regulatory framework with a view to identifying any gaps or anomalies and setting out a regulatory framework that addresses these.

– The commissioner should examine and assess the potential impact of biometric technologies on the privacy of individuals, the security of their information and the quality of the services they can be offered and that the public can access.

This will require a thorough examination of the use of biometrics, including the use of biometrics in access controls, biometric-enabled applications in financial transactions, biometric identity documents and biometric identity cards, including cards that are used by government agencies.

Identity

An identity commissioner could develop appropriate powers to investigate matters and to issue a final order on a wide range of matters, including orders that the commissioner is satisfied are required to protect the privacy of individuals and the security of their personal information, including orders

;;;;;;;;under the Data Protection Act 1998.

11. The necessity of appointing an identity commissioner

It is past time to appoint an identity commissioner with appropriate authority and responsibilities.

The commissioner should conduct an independent examination of the current regulatory system in order to identify any flaws and set up a new mechanism to address them.

The responsibility of an identity commissioner sh
framework for any gaps or anomalies and to esta
them. This is an appropriate role.

;margin;margin;margin;margin;margin;margin;margin;margin

Note. Here is a remark.

A commissioner who is responsible for establishing a biometrics regulatory framework must consider the proper scope of the commissioner's powers. The identity commissioner should investigate and evaluate individual privacy, personal information security, and service quality. This person should be able to do so with the help of biometrics.

The commissioner will be able to investigate and evaluate:

- Legislation, guidelines, or other regulatory mechanisms are required to address the privacy concerns raised by biometrics.
- to address the security risks posed by biometrics use through legislation, advice, or other regulatory mechanisms
- Any type of regulation, guidance, or other regulatory framework that addresses biometric quality assurance and service quality is urgently needed.
- The use of biometrics must meet the commissioner's minimum standards for privacy, security, and quality assurance.

12. The current legal framework for biometrics

In the United Kingdom, there is already a legal framework for biometrics, but it was piecemeal and is not yet comprehensive.

Some of the framework's flaws are as follows:

This country does not appear to have a comprehensive biometrics framework in place.

In the United Kingdom, biometrics for identification documents are not governed by a comprehensive legislative framework.

Financial entities' use of biometric identification is not clearly regulated.

There is no well-defined Knuth [3] regulatory framework for biometric identification by government entities.

There is no clear regulatory structure for biometric access controls.

Europe's role in establishing a strong regulatory framework

As a global leader, the European Union has created a strong regulatory framework.

Grandstrand [2]

Acknowledgments

I would like to thank ...

References

J. Fagerberg, D.C. Mowery, and R.R. Nelson (Eds.). 2004. Oxford Handbook of Innovation. Vol. 1. Oxford University Press, Oxford.

O. Grandstrand. 2004. Innovation and Intellectual Property Rights, Chapter 10. Volume 1 of Fagerberg et al. Fagerberg, Mowery, and Nelson [1].

Donald E. Knuth. 1984. *The TeX book*. Addison-Wesley.

A. An appendix

“All the same, I should like it all plain and clear,” said he obstinately, putting on his business manner (usually reserved for people who tried to borrow money off him), and doing his best to appear wise and prudent and professional and live up to Gandalf’s recommendation. “Also I should like to know about risks, out-of-pocket expenses, time required and remuneration, and so forth” – by which he meant: “What am I going to get out of it? and am I going to come back alive?”

B. Conclusion

Really fun to write Markdown :-)