INTERNATIONAL LABOUR ORGANIZATION

Seafarers’ Identity Documents Convention (Revised), 2003 (No. 185)

ILO Seafarers' Identity Documents Biometric Testing Campaign Report

(Addendum to Part I)

Geneva, 2005

INTERNATIONAL LABOUR OFFICE GENEVA

Foreword

The International Labour Organization (ILO), established in 1919, is a Specialized Agency of the United Nations (UN). It is a tripartite organization, in which representatives of Governments, Employers and Workers take part with equal status. In June 2003, the ILO adopted the Seafarers' Identity Documents Convention (Revised), 2003 (Convention No. 185). The revision of the earlier Convention of 1958 was prompted by discussions held in the International Maritime Organization (IMO) reviewing measures and procedures to prevent acts of terrorism that threaten the security of passengers and crews and the safety of ships. ILO Convention No. 185, which came into force on February 9, 2005, is a binding international treaty for all Members that ratify it.

Implementation of ILO Convention No. 185 will represent the world's first internationally interoperable biometric verification system. In March 2004, the ILO Governing Body adopted the technical standard, ILO SID-0002 Finger Minutiae-Based Biometric Profile for the Seafarers' Identity Documents, which will be used to enable global biometric interoperability of Members' implemented systems (as specified in ILO Convention No. 185). The biometric storage format described in ILO SID-0002 was based on draft ISO standards dated October 2003, but minor modifications were made to satisfy the requirements of storing two fingerprint templates on a two-dimensional PDF417 barcode. Since the ISO standards were still in a relatively early draft form, no manufacturers were known to have products that supported these standards. Consequently, modifications to commercial products were necessary. In order to ensure that products supporting these standards, particularly the draft version of ISO 19794-2 specified in ILO SID-0002, could provide adequate interoperable performance on real seafarers, the ILO commissioned the ILO SID Biometric Testing Campaign to develop a list of compliant biometric products for Members to use when implementing ILO Convention No. 185.

The ILO Biometric Testing Campaign consisted of two phases. In the first phase, several biometric algorithm and sensor pairs (referred to as biometric products) underwent preliminary evaluation to determine those systems with sufficient conformance to the standards and basic matching performance to be included in the second phase of testing. Seven products were included in the second phase, which was conducted onboard a seafaring vessel. The experimental procedures, results, and analysis were included in the document, ILO Seafarers' Identity Documents Biometric Testing Campaign Report - Part 1, wherein the tested systems are referred to as Products A through G.

Only two of the seven products, A and F, achieved the ILO targets for both native and interoperable performance, and so it became apparent that interoperability using the standard might not be as simple as had been anticipated. A follow-on study was commissioned to investigate what had caused the problems in interoperability, with the eventual hope of performing additional testing with the same group of products to determine if interoperability could be enhanced. That study is the subject of this report.

Executive Summary

In 2004, the International Labour Organization (ILO) commissioned a Biometric Testing Campaign to evaluate biometric products for use in a globally interoperable system of Seafarers' Identity Documents. This test took place during September and October, 2004 on board the cruise ship Crystal Harmony, with live fingerprint data collected on seven different biometric products (each consisting of a sensor and algorithm pair) from 126 seafarers. The test was designed to simulate real operations, so genuine, imposter and interoperability testing were all done using live captured images, leading to a maximum of 210 fingerprint captures (principally of the right and left index fingers) for each seafarer during verification, plus an additional 14 for enrollment. During enrollment, all of the biometric products had to produce a BioAPI Biometric Interchange Record (BIR) in the form specified by the technical document ILO SID-0002. This BIR contained fingerprint minutiae records for the two fingers (usually the left and right index finger) enrolled for that seafarer, with the records in the form of standardized minutiae records based on the card normal format specified in the draft of ISO 19794-2 dated 2003-10-07.

Based on live testing with both genuine and imposter seafarers, every product except one (which appeared to have image acquisition problems and thus was not relevant in predicting performance) achieved a false reject rate of 5% or better at a false accept rate of 2% on the basis of a match being declared if either of the seafarer's two fingers could be verified within three attempts each. Given the uncertainty of 1-2% in the measurements, this was reasonable. When BIRs produced on one product were used to try and match against a seafarer placing their finger on another product, however, the false reject rates varied from 0% to 56%. Clearly some pairs of products were interoperable and others were not. At the ILO's desired performance goal of 1% FAR and 1% FRR, only 2 of the 7 products were interoperable, and these two products were the only ones promoted to the ILO list of approved products for SID use. This raised the concern that vendors implementing the fingerprint standard specified in ILO SID-0002 might not easily achieve interoperability, possibly due to ambiguities in the standard or areas that might be difficult to interpret correctly. The ILO decided to commission a follow-on test. The goal was to analyze the possible sources of interoperability problems, attempt to have the vendors provide new versions of the enrollment and matching algorithm components of their products based on this analysis, and then test these in an offline test using the images stored during the original ILO test. This would effectively provide an estimate of how well the products would have performed in the original test if the vendors had been aware of all the interoperability problems at the time they originally submitted their products.

Although the ISO 19794-2 version specified in ILO SID-0002 was a draft standard, it has now become a final draft international standard (FDIS) and the components related to the card normal format have not changed in any significant way. This means that the results of the ILO test were expected to have significance for other systems attempting to use an interoperable fingerprint minutiae record. Due to this interest, financial support was provided for the new ILO test by the US based National Biometric Security Project.

After extensive discussions with the participating vendors and a review of the possible sources of ambiguity in the standard, the product vendors also had the opportunity to analyze each others' BIRs. Based on all of this, new software was provided and a second phase of testing conducted with six of the original products. In this test, all of the products achieved a false reject rate of 3% or better at a false accept rate of 1% using their own BIRs. This was a slight improvement over the previous test, but within the margin of error. During interoperability testing, however, the worst case false reject rate was only 5% instead of 56%, indicating that all the significant interoperability problems had been resolved. In order to compensate for the fact that errors caused by failure to acquire a usable image from a fingerprint were not possible in a test using previously saved images, the failure to acquire rates (FTAR) for the various products in the original test were added to their FRR values in this test. At the ILO desired performance goal of 1% FAR and 1% FRR, the new test indicated that up to 3 products could be declared interoperable.

It thus seems prudent that the ILO list of approved interoperable products at this time be expanded to include three products. Specifically, Products A and F (the original interoperable pair) are now additionally interoperable with Product C, based on performance during the "Key Visit", when operator instruction was permitted to assist the seafarer in placing their fingers on the sensors. The maximum FRR in any interoperable combination of these four products during the Key Visit was 1.6%, even including the effects of FTAR as measured in the original test.

The majority of the problems rectified by the vendors seemed to relate to issues in implementing the standard. Hopefully that can be addressed by the final draft of a guidance document intended for anyone attempting to implement any standard related to ISO 19794-2 that has been included in Annex B of this document. Three of the issues are areas where there are clear ambiguities in the underlying ISO 19794-2 standard and these should be addressed both in an ISO amendment to the standard and in future revisions of ILO SID-0002. The first is that minutiae type in the standard may be "ridge ending", "ridge bifurcation" or "other". Some products use "other" to encode endpoints or bifurcations where they are unsure of or don't care about the minutiae type. Other products require the minutiae type to be correctly identified as "ridge ending" or "ridge bifurcation" and will ignore "other" minutiae. The standard should be amended to ensure that any matching algorithm claiming the ability to use standardized data interchange records must be able to accept records containing "other" type minutiae. The second major issue is that the minutiae angle can be computed in different ways and some products end up with a quantized angle for each minutiae point. Other products may anticipate that the angle is quantized to the encoding quantization level specified in the standard and may set internal tolerance parameters based on that. The standard should thus be amended or a separate guidance document produced to indicate that in calculating minutiae angles, they should be quantized as closely as possible to the encoding quantization level defined in the standard. The third issue was that the method of truncation of minutiae specified in the standard was considered by all vendors to be unclear. In fact, most vendors commented that they use their own method of truncation, presumably because the standard allows the leeway of truncating based on minutiae quality. A clearer method of truncation should be specified in any future amendment to the standard. One suggested method is included in Annex B of this report.

Overall, the new versions of the vendor software had no significant interoperability problems, indicating that an amended version of the card normal format of ISO 19794-2, as could be specified in any future versions of ILO SID-0002 is capable of providing adequate matching performance and interoperability, provided the vendors implement correctly. Annex B of this report should be provided to any vendors attempting to produce systems for use with ILO SID applications in future, both so that they can understand the general implementation guidance, which may help them avoid some common pitfalls, and so that they can be aware of the issues that are currently ambiguous in both ISO 19794-2 and ILO SID-0002, and of how they should resolve them. Specifically, matchers must be able to match even if minutiae are all of "other" type, minutiae encoders must attempt to estimate angle to the angular quantization specified in the standard and truncation after all low quality minutiae have been eliminated can be simplified by using the method in Annex B.

Contents

Index of Tables

Index of Figures

1. Introduction

1.1. Background

Fingerprint matchers based on minutiae points have been a mainstay of the biometrics industry since its infancy. In the past, all of the different vendors typically used a proprietary format, commonly known as a proprietary template, to store the sets of minutiae and associated features they extracted and used for matching. This prevented interoperability, since a template created by a product from one vendor could not be correctly interpreted by other vendors' products and thus enrollment and verification had to use equipment supplied by the same vendor. One way to resolve this problem was to exchange the entire fingerprint image, but this was not always possible, due to storage, network bandwidth or privacy issues. Recently, work began to determine the basic elements comprising minutiae templates and formalize them in a standardized minutiae template. This was first accomplished in ANSI INCITS 378:2004 Information Technology – Finger Minutiae Format for Data Interchange, a US standard published in 2004. At the same time, the International Organization for Standardization (ISO) was developing an international minutiae standard through the ISO/IEC JTC 1 SC 37 group. This standard has not yet been published.

In June, 2003, The International Labour Organization, a specialized agency of the United Nations, adopted the Seafarers' Identity Documents Convention (Revised), 2003 (Convention No. 185). This Convention defines a globally interoperable system of Seafarers' Identity Documents that will be used to verify that seafarers around the world have a right to the special privileges related to transit through or entry into countries (such as for shore leave) that the Convention grants them. Convention No. 185 determined that the seafarers' identities would be verified using a biometric stored in a 2-D bar code on the document. Given the limited storage capacity of the 2-D bar code, a template was the only choice, and in 2004, the ILO approved ILO SID-0002 Finger Minutiae-Based Biometric Profile for the Seafarers' Identity Documents. This document defined the standard for the biometric template to be stored in the bar code and the method for enrollment and verification of the seafarers' fingerprints. The minutiae format selected was a draft of the SC 37 standard ISO 19794-2 Biometric Data Interchange Formats – Part 2: Finger Minutiae Data dated October, 2003 and the specific format selected was the normal sized finger minutiae card format ("card normal format").

Since the standard was still in a draft format, and since nobody had ever deployed a globally interoperable biometric system using standardized templates, there was a significant risk that seafarers who were enrolled in their home country as they received their SIDs might have difficulty being verified by equipment from a different vendor at a port in another country. The ILO therefore decided to conduct a Biometric Technology Test using a real population of seafarers on a ship, with multiple biometric products (each consisting of a fingerprint sensor combined with an enrollment and verification algorithm) from different vendors, to determine whether or not vendor claims of adherence to the standard could produce interoperability with an acceptable level of biometric matching performance, as measured by false reject rate (FRR) at a fixed false accept rate (FAR). The target was to achieve a 1% or better FRR at a 1% FAR. A call went out to vendors, and those who chose to respond provided biometric products for preliminary testing. This included verification that the basic structure and format of the output data from this software was a minutiae based interchange format in conformance to ILO SID-0002. Seven products were declared conformant and thus suitable for the full test, which took place in September and October of 2004. In order to simulate operational enrollment and verification of seafarers in a realistic environment, all the tests involved live capture of fingerprints from seafarers. The detailed test methodology employed is described in Section 5 of the ILO Seafarers' Identity Documents Biometric Testing Campaign Report document. With multiple products and the requirement of live capture, time restrictions meant that only 126 seafarers participated in the test, but each of them enrolled two fingers on each product and then attempted to verify multiple times on each product, resulting in a total of 26,067 live finger verification attempts.

The results of this test were easy to summarize. Every product except one (which appeared to have a sensor problem and thus was not relevant in predicting performance) achieved a false reject rate of 5% or better at a false accept rate of 2% on the basis of a match being declared if either of the seafarer's two fingers could be verified within three attempts. Given the uncertainty of 1-2% in the measurements, this was reasonable. When BIRs produced on one product were used to try and match against a seafarer placing their finger on another product, however, the false reject rates varied from 0% to 56%. Clearly some pairs of products were interoperable and others were not. Only 2 of the 7 products met the ILO's desired performance goal of 1% or better FRR at a 1% FRR. All of these results and the methodology used for the test are clearly described in the ILO Seafarers' Identity Documents Biometric Testing Campaign Report

In a gesture of cooperation between ISO and ILO, ISO gave permission for ILO to include the complete text of the draft 19794-2 standard in the ILO SID-0002 document. The part of 19794-2 that describes the card normal format has not changed significantly since that draft, and since this standard is now at an FDIS stage, it is unlikely to change. The card normal format is also quite similar to the basic information in the ANSI INCITS 378:2004 standard. This meant that the results of the ILO test were quite relevant for other groups around the world who intended to deploy systems based on an interoperable fingerprint minutiae template. In fact, the ILO was the first to test such a standardized template among multiple products from multiple vendors, and as such, this initial report was groundbreaking in its scope.

1.2. Current Test

When, the results of the ILO investigation, and the possible ramifications for other users of fingerprint minutiae standards were revealed, there was some concern, since the standard defining the fingerprint minutiae was the most likely source of ambiguities leading to a lack of interoperability. Since this was a draft ISO standard, ISO/IEC JTC 1 SC 37 offered a list of international experts to help with understanding and resolving any ambiguities. These experts have, in fact, reviewed this report and their input has helped to make it a technically stronger report.

After considering the situation, the best solution was deemed to be a second test that would use the database of images collected during the live test on board the ship. ILO agreed that the test could be useful for potentially expanding the list of products meeting ILO requirements for use in SID application and also that resolving the interoperability issues would help in the long term with SID applications. Thus, a second test was commissioned.

In this second test, the goal was to repeat the original test in an offline mode using the images stored from the data collection on board the ship. Additionally, since the test was being done offline, it would now be possible to directly test imposters whose images came from one product against BIRs of other seafarers enrolled on different products. This imposter testing was skipped in the interoperability portion of the original test, and the imposter distributions from the same product were used to estimate an appropriate threshold score for all verifications on that product, regardless of which product was used for creating the enrollment BIR. This was simply due to the time constraints involved in a live test, which were not an issue in the offline test.

By using images from the original test and paying careful attention to the methodology originally employed, the goal was to simulate the results of the first test by emulating the sensor portion of each product with the stored images. Then the vendors would be free to revise the software portion of their products (enrollment and matching algorithms) to try and improve interoperability and the test could be repeated to estimate what the results would have been if the interoperability issues had been understood and the updated software available during the original ship board test. The detailed methodology surrounding this process is described in Section 2.

It was also necessary to carefully analyze the minutiae standard ISO 19794-2 itself, and especially the way it had been implemented by the vendors in accordance with ILO SID-0002 to see where the areas of ambiguity or confusion were that might have caused the interoperability problems. In this area, the ISO experts were particularly willing to help, and SC 37 provided comments to ILO on some possible problems. These were discussed among many standards experts with feedback from various groups, until eventually the document presented in Annex A was prepared by the US based National Biometrics Security Project as a technical contribution to interested biometrics standards bodies and implementers (including ILO) to help them in revising their standards. Since this seemed very appropriate for this test, it was distributed to the vendors of the products involved in the ILO Biometrics Testing Campaign for them to consider as they attempted to resolve interoperability issues in their software. Later, based on comments from other participants in the standards process, the revised document contained in Annex B was produced. This was not sent to the vendors, as it was produced too late, but it is included here for completeness.

All of the vendors were asked to provide two new versions of the software component of their products. The first was to use exactly the same enrollment and matching algorithm as was used in the original test, but modified for offline testing. This was used in an offline "*baseline*" test to ensure that the baseline results obtained in the offline test were comparable to those obtained in the original test, thus ensuring that the methodology used to simulate the original results was valid. The second version of the software was to incorporate any changes that the vendors felt were necessary to improve interoperability after reviewing the guidance document sent out. It was used in a second offline test referred to as the "*enhanced*" test which provided results that reflected any improvements to interoperability resulting from the changes.

The vendor of product "D" from the original test decided that their sensors were the source of their problems and thus there was nothing they could do by changing software. They declined to participate. Vendors of the other six products sent modified software, although Vendor "A" sent only one version of its software. Based on the test results, it appears that this version was identical to that used in the original test, indicating that Vendor "A" did not feel that any changes to their software were necessary. Thus, the same software was used for vendor "A" in both the baseline test and the enhanced test.

The results of both the baseline and enhanced tests for all six participating products are contained in Sections 3 and 4 and in Annexes C and D. The final conclusions are in Section 5.

2. Test Methodology

The methodology used for all three of the primary components of the original test; Conformance, Performance, and Interoperability testing, are clearly described in section 5 of the ILO Seafarers' Identity Documents Biometric Testing Campaign Report.

The methodology used here tested the same three fundamental components, but using a specially written software tool to automate the testing. There were some functions in this tool that were set-up by the user at the beginning of each test session, but most of the test time was simply spent by the tool accessing the database and running enrollment or match operations. It should be noted that the conformance step during the enrollment phase tested all of the same things about the Biometric Interchange Records produced as were tested in the original test and these are described explicitly in Section 5.1.1 of ILO Seafarers' Identity Documents Biometric Testing Campaign Report. This was done just in case one of the products had been incorrectly implemented when being converted for offline testing or being modified for enhanced interoperability. It turned out that no conformance problems were found, so this was not an issue for any of the products. It should also be noted that the images collected during seafarer enrollment and in each of their subsequent visits were addressed separately in the offline test so that results could be extracted separately for each visit, for all visits counted together, or just for visit 3, the "Key Visit", exactly as was defined in the original test.

It is important to remember that each seafarer in the original test was enrolled using both a primary and secondary finger, preferably left and right index fingers into a two-finger BIR. All matching tests followed the procedure outlined in ILO SID-0002 in that a match was considered to have taken place if the seafarer could match against the enrollment BIR in a maximum of three placements of each of the primary and secondary finger. Since placements were not applicable in the offline test, but multiple images were available for most seafarers' primary and secondary fingers due to the large number of verification tests conducted in the original test, the ultimate match score used in computing all performance metrics was the best match (in BioAPI terms the minimum FARAchieved value) in up to three match attempts for each of the primary and secondary finger. If three images were not available, then all available images were used and the best match score taken. This meant that achieving a 1% FAR was more difficult but achieving a 1% FRR was easier than it might be in other offline tests where only a single match attempt for a single finger is considered in producing the final score.

The offline test was divided into an enrollment phase and a verification or matching phase, and these are described below using a simple point form outline approach coupled with a block diagram. Each vendor had provided software according to a very simple offline testing software description that is included as Option 2 in the "Vendor Software Requirements" document included in Annex E.

This document was sent out to the vendors and after some discussions with them, Option 2 was selected as being simpler to implement.

2.1. Enrolment Phase

User chooses Test (Baseline or Enhanced)

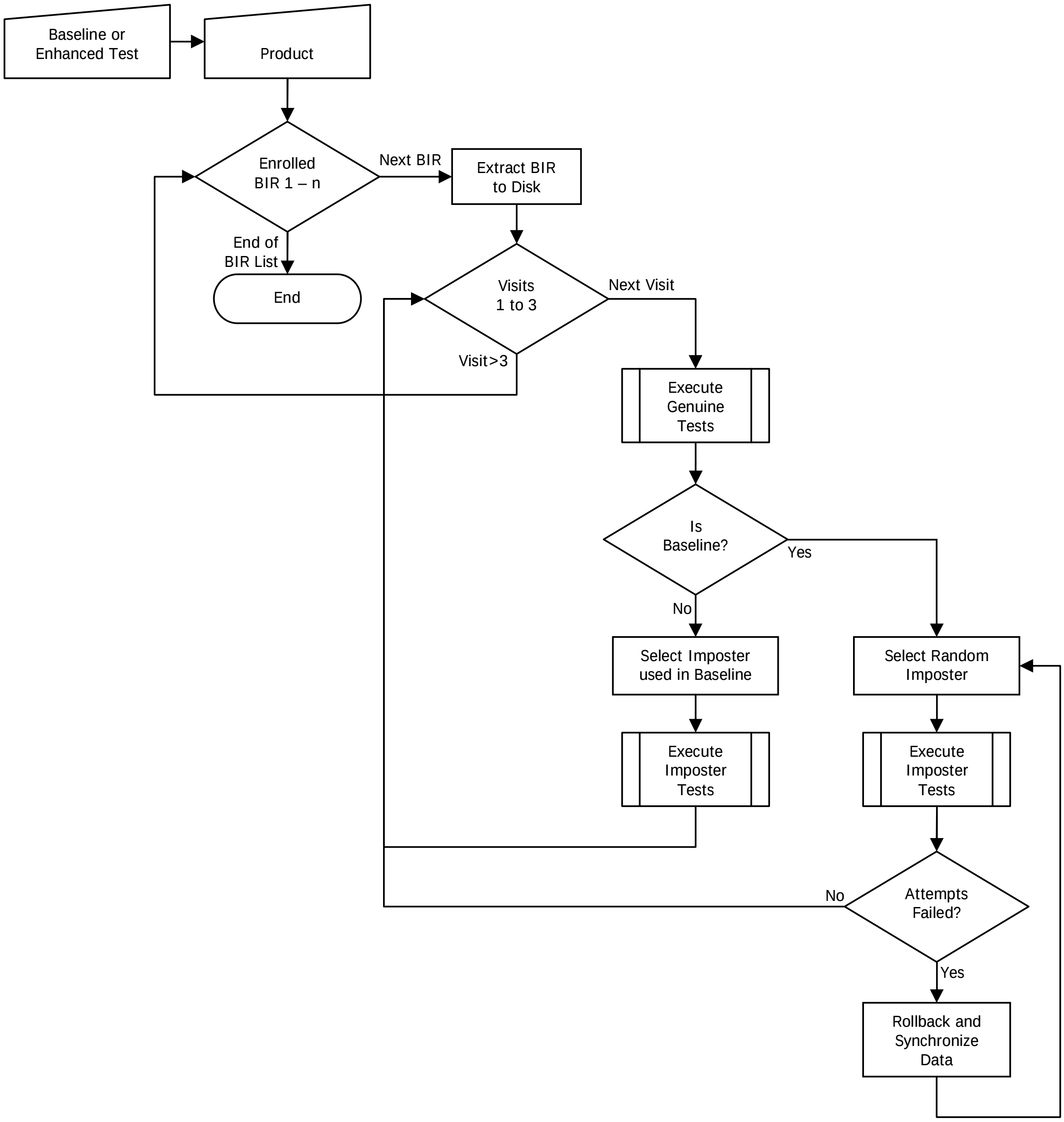
User chooses Product

For each seafarer record containing images of both primary and secondary fingers stored during the original live test:

- Extract images to disk.
- Run product's (selected in step 1) associated sidenroll.exe developed for the current test.
- If BIR file exists, check conformance to SID-0002 and store in database.

2.2. Verification Phase

2.2.1. Verification Process Flowchart



2.2.2. Genuine Tests

The following steps are run once for the primary finger from each enrolled BIR and repeated again for the secondary finger to verify up to three (3) simulated attempts per finger (max. six attempts per BIR) for each visit on the product currently being tested:

- Retrieve up to three (3) genuine image records for the current product, visit, and seafarer which correspond to the primary/secondary finger of the current BIR.[*]
- For each image retrieved:
  - Extract image to disk.
  - Run product's associated sidverify.exe developed for the current test.
  - If the subsequent match/score result file exists, store these values to the database.[†]

[*] In the event that there are more than three images available, only the most recently acquired images of the set are retrieved. If no images are available for a particular finger, no genuine verifications can be attempted for that template's finger, product, and visit combination.

[†] These simulated two-finger, three attempt transactions are stored as such within the database to enable FRR and FAR scores to be computed on a transactional basis. Each two-finger score is the result of finding the minimum score of each single-finger transaction. The score for a single-finger transaction is therefore, the minimum score of each verification "attempt" (up to three per finger). If no verification attempts can run for one of the transaction's two fingers, the minimum score of the single-finger transaction is used.

2.2.3. Imposter Tests

During the first baseline test, each imposter is selected at random (see Select Random Imposter). The enhanced test however, uses the imposters that were selected during the baseline test in order to ensure that exactly the same tests are being re-run with the enhanced software.

The following steps are run once for the primary finger from each enrolled BIR and repeated again for the secondary finger to verify up to three (3) simulated attempts per finger (max. six attempts per BIR) for each visit on the product currently being tested:

- Retrieve up to three (3) imposter image records for the current product and visit which corresponds to the primary/secondary finger of the current BIR.[‡]
- For each image retrieved:
  - Extract image to disk.
  - Run product's associated sidverify.exe developed for the current test.
  - If the subsequent match/score result file exists, store these values to the database.[§]

Select Random Imposter (Baseline Test)

To select an imposter to be used in the subsequent Imposter Test process, the database is queried for a list of all seafarers that have at least one image corresponding to each of the primary and secondary fingers in the BIR for the current product and visit. Any imposters that may have already failed all tests (see "Attempts Failed?" below) on a particular finger for this BIR and visit on the current product are filtered out.

If no records can be found to meet the preceding criteria, no imposter verifications are executed for this BIR, product, and visit combination. If one or more records are returned however, the test software randomly chooses one of these imposter seafarers to be used.

[‡] In the event that there are more than three images available, only the most recently acquired images of the set are retrieved.

[§] These simulated two-finger, three attempt transactions are stored as such within the database to enable FRR and FAR scores to be computed on a transactional basis. Each two-finger score is the result of finding the minimum score of each single-finger transaction. The score for a single-finger transaction is therefore, the minimum score of each verification "attempt" (up to three per finger). If no verification attempts can run for one of the transaction's two fingers, the minimum score of the single-finger transaction is used.

Attempts Failed?

For the baseline test, a test attempt is said to have failed when a product's baseline sidverify.exe fails to return a match/score result file. This primarily occurred because of random errors experienced by some products in the original test where they returned a fingerprint image file to disk that was blank or contained only a partial fingerprint, as opposed to the normal image they displayed on the screen. Since the original test operators were viewing the images displayed by each BSP on the screen, this problem was not noted until the offline test was underway. It is probably related to the fact that some BSPs use a stream of images over time to eventually settle on the image they will use, but because the function to save an image to disk was specially added for the test and not part of the normal BSP operation, it may not have used the final image, but one of the interim ones when the sensor was first triggered or before the appropriate gain changes had been made. This meant the database had some fingerprint image files with no usable content, which caused certain products to halt without producing any output. Since this was an artifact of the artificial nature of image capture during the original test, it was decided not to penalize products for failure to enroll or acquire with these images. It was still necessary to continue the test, however, so each time this occurs, a counter is incremented. If all test attempts fail on either the primary or secondary finger for the current BIR and visit, the test software will try to select an alternate imposter to repeat the two-finger transaction.

Rollback and Synchronize Data

If a product's baseline sidverify.exe failed to return a match/score result file on all tests for a particular finger for each BIR and visit;

Any successful tests on the primary finger are removed from the database.

The current imposter's SeafarerID is added to a temporary list for this BIR and visit to prevent this imposter from being selected again. (see Select Random Imposter)

3. Test Results and Data Analysis

This section presents a summary of the most important raw results. It is critical for an understanding of how the final conclusions were achieved.

3.1. Summary of Caveats

Before beginning to consider the results themselves, it is important to recognize that although this test is a best effort to simulate the results that would have been achieved on board the ship in the original test, there are some minor issues that can't be accounted for. None of these were expected to be significant in comparison to the magnitude of the interoperability problems being reviewed, but they are listed here for completeness.

1. In the original test, there were no imposter attempts in the interoperability section of the test. The threshold established for a FAR of 1% or 2% (depending on the results being reported) in the test using genuine and imposter seafarers enrolled on the same product was also used in the tests for genuine seafarers being verified against their enrollment templates for other products. It was expected that this may have made the FRR reported in the original test slightly higher than it would otherwise have been. The fact that a complete set of imposter tests were performed in the offline test may make it more accurate than the original test, but may have marginally reduced the overall FRR for every product.
2. In the original test, there was a random selection of imposter seafarers in the native performance portion of the test and also a randomly selected subset of genuine seafarer BIRs in the interoperability portion. In this test, the random imposter selection was different than in the original test and the interoperability tests were complete. Thus, the selection of fingers being compared in the verification portion of the test was different than in the original test. Provided both tests included a sufficient number of comparisons, this should not have changed any individual result by more than twice the margin of error of 1-2%.
3. In the original test, seafarers made up to three placements of their fingers in each attempt to match, with GUI feedback from the BSP. In this test, the single image output by the BSP from each match attempt and stored in the database was used as if it represented a single placement. This may have made the image more likely to match than a single placement attempt (although this depends on how the BSP generated the output image), but the lack of user feedback from the BSP between placements would make the images less likely to match. Overall, it is difficult to quantify how much impact this may have had on the test, but it is a fundamental limitation of an offline test.
4. Since all of the images had to be successfully captured before they could be stored in the database, this offline test eliminated any errors arising

from failure to acquire a usable image. This was compensated for, however, by adding the failure to acquire rates from the original tests to the false reject rates recorded in this test. In fact, the correct formulation to account for failure to acquire tests also includes a second order effect involving false non match rate multiplied by failure to acquire rate. That term was zero in all cases for the key visit, and had a maximum value less than 0.05% for all visits combined, so it was deemed small enough to ignore in these tests.

5. The fact that some images were not properly output by the BSPs in the original test meant that only a subset of images could be analyzed in the offline test. Visual analysis of these images and their locations in the database suggested that this was a random event, but it is possible that not all of the nominally "bad" images were actually due to failures to correctly record the image in the original test. This means that there could have potentially been some failures to enroll or failures to acquire in this test that were attributed to improperly captured images, but were actually due to the image quality criteria being changed between the original software and the new offline software. Unfortunately there was no way of resolving this other than manually reviewing all the images and having a human determine if the product should have been expected to enroll or acquire each nominally "bad" image. This was done, and it showed that most of the images recovered from the database in this category were either blank, featured a partial print, or were in some other way obviously unacceptable. This was a highly qualitative analysis, however, and some images seemed to be on the borderline where enrollment or matching might have reasonably been expected to occur. There was simply no way to quantify this. Only one seafarer had a failure to enroll on one system during the entire original test. Also, the enrollment procedure in the original test allowed multiple placements from multiple fingers to find a pair that could be enrolled, whereas the offline enrollment only used one image from each of the two originally enrolled fingers (since that was all that was available). Overall, the circumstances around enrollment and image acquisition in the offline test are different from the original test and so it seems fair in trying to simulate the original test to ignore failure to enroll (as was done in the original report because only 1 seafarer failed in all enrollments on all systems) and to use the failure to acquire rates measured in the original test. This should give the closest simulation to the original results possible in this type of offline test.

3.2. Original Test

The first results to be considered are those from the original test that took place on board the ship. The results for product "D" have been removed from all tables, because that product did not participate in the offline tests. The results shown below in Table 1 are for the "Key Visit" where test operators provided feedback and assistance to people having trouble with finger placement. The key visit was also the last visit, which took place about 4 weeks after the enrollment. The results shown in Table 2 are for all three visits, the first of which occurred immediately after enrollment, the second about two weeks after enrollment, and the third being the key visit.

As in all of the other tables showing FRR among the six products, the diagonals highlighted in gray represent the performance of each product when verifying seafarers using BIRs created during enrollment on that same product (the native performance). The off-diagonal elements represent the interoperable performance.

It is obvious that some products performed quite well alone. A, C, F and G, for instance, all had native performance better than 2.5% FRR at 1% FAR in the key visit and in the average of all visits. Some of the interoperable pairs also performed well. Enrol on A and verify on C or F were both reasonable, as was enrol on anything and verify on A. Other pairs caused major problems, however. The most notable were Enrol on A or F and verify on B, E or G, as well as enroll on C and verify on F and enroll on G and verify on C.

Table 1. Original Live Test: Dual-Finger FRR at 1% FAR for Key Visit

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	72.6%	3.6%	52.0%	1.6%	40.6%
B	4.7%	100%*	7.1%	12.1%	4.7%	4.5%
C	0.0%	9.4%	1.7%	5.5%	59.4%	3.0%
E	1.9%	14.1%	6.3%	4.9%	1.8%	3.6%
F	0.0%	58.3%	4.9%	41.9%	0.0%	27.3%
G	4.3%	7.8%	46.6%	6.3%	17.0%	1.6%

* Product B was unable to achieve a 1% FAR, so the FRR was artificially set to 100%.

Table 2. Original Live Test: Dual-Finger FRR at 1% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	1.9%	100%*	7.0%	41.7%	1.4%	49.6%
B	3.2%	100%*	7.8%	5.6%	5.8%	5.1%
C	0.0%	100%*	2.4%	2.5%	65.0%	4.8%
E	4.1%	100%*	5.8%	3.2%	2.8%	2.4%
F	3.0%	100%*	5.6%	31.5%	0.3%	36.9%
G	3.6%	100%*	40.7%	9.4%	17.5%	1.9%

* Product B was unable to achieve a 1% FAR, so the FRR was artificially set to 100%.

3.3. Baseline Offline Test

The first offline test was run using the baseline version of the new software provided for the six products. FRR values were computed for the key visit and for all visits at 1% FAR. The FTA values from the original test (for the key visit and for all visits respectively) were added to these FRR values and then the corresponding FRR values from the original test were subtracted. If the original test and the offline test with the original versions of the algorithms are equivalent, then both of the tables should contain all zeros. Given the caveats discussed above, we expect them to be between -5% and +5%, with a slight bias towards the positive numbers.

Table 3. Difference (Original - Baseline Test): Dual-Finger FRR at 1% FAR for Key Visit

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	35.6%	2.7%	18.4%	0.7%	-1.9%
B	2.9%	98.3%	3.6%	100%*	-0.6%	-2.6%
C	0.0%	7.0%	2.0%	1.2%	-3.5%	-3.9%
E	0.9%	8.4%	4.4%	-1.7%	-3.0%	-1.1%
F	-0.9%	38.5%	5.6%	20.4%	0.0%	-4.5%
G	2.2%	-0.3%	44.5%	-6.0%	15.0%	-3.8%

* Product E could not achieve a 1% FAR with Product B templates in the Baseline test for the Key Visit

Table 4. Difference (Original - Baseline Test): Dual-Finger FRR at 1% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	3.5%	52.3%	10.4%	6.8%	-1.1%	6.3%
B	2.2%	97.4%	5.9%	-0.5%	1.8%	-0.9%
C	0.0%	96.1%	3.6%	-1.4%	0.2%	0.2%
E	4.6%	96.3%	3.8%	0.2%	-1.2%	-2.5%
F	5.7%	74.7%	6.9%	8.2%	0.0%	3.0%
G	1.7%	92.7%	38.7%	-3.7%	15.5%	-4.8%

In fact, most of the entries in these two tables do fall within the 5% threshold, suggesting that the two tests are generally equivalent. Most of the exceptions are for Verifications on Product B or Product E. The reason is probably that these two products were working very close to their FAR limits, since in at least some of the tests in either the original test (for Product B) or the baseline test (for Product E) they could not achieve a FAR of 1%. Typically, because of the shape of the DET curve, the uncertainty in results from a product increases substantially when it is operating close to its FAR limit.

The other anomalies in the tables are for enrollment on Product G and verification on Products C and F. There was no immediately apparent reason for this, but after discussions with the vendor of Product G, they realized that they had made a mistake and included one of their changes to the software for the enhanced test in the baseline software. This explained these unexpected results.

Since the results for most of the entries in the tables were within expected tolerances, and since there were explanations for those that were not, it appears that the offline test does an acceptable job of simulating the original test.

A second baseline test, referred to as "Baseline Test #2" or just "Baseline 2", with a different random selection of imposters, was also executed to see if that made significant differences. The differences were once again within a few percent in most cases, as can be seen by reviewing the detailed results provided in Annex C and Annex D. This result again confirmed the validity of the offline test as a way of emulating the original test.

3.4. Enhanced Offline Test

The next test used the second version of the software which had been enhanced by each vendor to try and resolve their interoperability issues, based on the guidance provided in the document sent to vendors and on a detailed review of other vendors BIRs. The results, shown in Tables 5 and 6, were impressive.

Table 5. Enhanced Test: Dual-Finger FRR at 1% FAR for Key Visit

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	4.3%	0.9%	3.5%	0.0%	2.6%
B	2.7%	2.7%	2.7%	3.5%	1.8%	2.7%
C	0.0%	2.6%	0.0%	3.5%	0.9%	1.8%
E	1.0%	1.9%	1.0%	2.9%	1.0%	1.0%
F	0.9%	5.2%	0.0%	3.3%	0.0%	3.5%
G	0.9%	2.7%	2.7%	2.7%	0.9%	2.7%

Table 6. Enhanced Test: Dual-Finger FRR at 1% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	3.4%	0.9%	2.8%	0.3%	2.6%
B	2.0%	1.2%	3.2%	3.2%	1.4%	2.6%
C	0.3%	3.2%	0.9%	2.9%	0.3%	2.6%
E	0.6%	1.2%	2.5%	1.6%	0.6%	0.9%
F	0.3%	2.8%	0.6%	2.8%	0.0%	4.0%
G	0.6%	2.3%	2.3%	2.6%	0.9%	2.3%

There are now no entries in either table with excessive FRR values. In comparison to Table 1, with a maximum FRR of 59.4% (excluding Product B which was unstable because it was so close to its FAR limit), Table 5 has a maximum FRR of 5.2%. In fact, the average of all the diagonal values in Table 1 is 1.4% and the average of all values in the Table is 1.9%. This suggests there is minimal or no loss in performance due to interoperability.

In Tables 7 and 8 below, the effects of failure to acquire have been added back in, to get a completely fair comparison with the original test. Even so, the maximum FRR for the key visit is 5.2% and for all visits, it is 5.1%. Clearly the enhancements have resolved all substantial interoperability problems and the numbers now mostly reflect the failure to acquire rate of the different sensors used, the uncertainty of the tests, and the limits of each product against a population of seafarers, who may be a fairly difficult demographic group for fingerprint products in general.

Table 7. Enhanced Test: Dual-Finger FRR+FTA at 1% FAR for Key Visit

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	4.3%	0.9%	3.5%	0.0%	2.6%
B	2.7%	2.7%	2.7%	3.5%	1.8%	2.7%
C	0.0%	2.6%	1.2%	3.5%	0.9%	1.8%
E	1.0%	1.9%	1.0%	2.9%	1.0%	1.0%
F	0.9%	5.2%	1.6%	3.3%	0.0%	3.5%
G	0.9%	2.7%	2.7%	2.7%	0.9%	2.7%

Table 8. Enhanced Test: Dual-Finger FRR+FTA at 1% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	1.6%	3.4%	5.1%	2.8%	0.3%	2.6%
B	3.4%	1.2%	4.7%	3.2%	1.4%	2.6%
C	0.3%	3.2%	2.6%	2.9%	0.3%	2.6%
E	2.7%	1.2%	3.0%	1.6%	0.6%	0.9%
F	3.3%	2.8%	2.7%	2.8%	0.0%	4.0%
G	1.0%	2.3%	3.7%	2.6%	0.9%	2.3%

Further information about the number of genuine and imposter verification tests for each interoperability combination in the key visit and in all visits, as well as detailed performance tables for a FAR of 2%, are given in Annex C. A complete set of DET curves are given in Annex D.

4. Interoperable Product Combinations

Following the methodology of the original report, it was necessary to determine the mean, maximum and minimum FRR at the various operating points for the key visit and for all visits combined, in order to determine which products (and how many products) could meet the ILO stated requirements of a 1% or better FRR at a 1% FAR both in native and interoperable usage. It should be noted that the effects of FTA were added when computing these numbers and that the mean includes both native and interoperable performance.

Table 9. Summary of Best Performing Product Combinations

	Key Visit		All Visits	
	1% FAR	2% FAR	1% FAR	2% FAR
Best Combinations of 2 Products				
Products	A-F	A-F	E-F	E-F
Mean FRR	0.2%	0.2%	1.3%	1.2%
Max. FRR	0.9%	0.9%	2.8%	2.5%
Min. FRR	0.0%	0.0%	0%	0.0%
Best Combinations of 3 Products				
Products	A-C-F	A-C-F	B-E-F	E-F-G
Mean FRR	0.6%	0.6%	1.6%	1.5%
Max. FRR	1.6%	1.6%	3.2%	2.5%
Min. FRR	0.0%	0.0%	0.0%	0.0%
Best Combinations of 4 Products				
Products	A-C-F-G	A-C-F-G	A-E-F-G	A-E-F-G
Mean FRR	1.3%	1.1%	1.9%	1.7%
Max. FRR	3.5%	3.5%	4.0%	3.4%
Min. FRR	0.0%	0.0%	0.0%	0.0%
Best Combinations of 5 Products				
Products	A-C-E-F-G	A-C-E-F-G	A-B-E-F-G	A-B-E-F-G
Mean FRR	1.6%	1.4%	2.1%	1.9%
Max. FRR	3.5%	3.5%	4.0%	3.4%
Min. FRR	0.0%	0.0%	0.0%	0.0%
Best Combinations of 6 Products				
Products	A-B-C-E-F-G	A-B-C-E-F-G	A-B-C-E-F-G	A-B-C-E-F-G
Mean FRR	2.0%	1.8%	2.3%	2.0%
Max. FRR	5.2%	4.4%	5.1%	4.8%
Min. FRR	0.0%	0.0%	0.0%	0.0%

It is apparent from Table 7 that during native verifications only Products A and F have a false reject rate of less than 1% at a false accept rate of 1% during the key visit. Combining all visits, which involves three times as many verifications and three times as many randomly selected imposters and thus may be more statistically accurate, only Product F achieves the desired native performance level. On the other hand, the worst case native performance FRR value was 2.9% for the key visit and 2.6% for all visits. Given the uncertainty in the measurements of around 2%, there is not a lot of difference among these single values.

It is thus most logical to use the mean FRR calculated over all native and interoperable performance values for a particular combination of products as the most accurate means of determining which products meet the ILO performance requirements, since that number should be more stable.

On that basis, it appears that the combination of Products A, C and F easily meets the ILO targets during the key visit when operator guidance was provided to the seafarers after their first poor placement. During other visits, when no guidance was given, Products A and C exhibited a substantial increase in Failure to Acquire errors, and there is not even a pair of interoperable products that meet the ILO performance requirements. Given that the uncertainty in the measurements is approximately 1-2%, it probably makes sense only to consider performance values to the nearest full percent. In that case, the combination of Products A, C, F and G could also be said to have satisfied the ILO requirements during the key visit, with a mean FRR of 1%. Given that there were no recorded failure to acquire errors with product G during the original test, the addition of Product G to the list of approved products would be appealing. Unfortunately, Product G was most affected by the presence of the nominally "bad" images in the offline tests, and only enrolled 101 of 119 seafarers present in the database. Visually inspecting the images that were not enrolled, some of them did appear to be reasonable images and not images that were obviously due to problems in storing the images in the original test. Thus, given that the combination of Products A, C and F clearly met the performance targets and that no cases of nominally "bad" images were noted with any of these products, it is more prudent for now to only qualify these three products as having met the ILO requirements.

5. Conclusions

The current offline test is a reasonable simulation of expected results if the enhanced software to resolve interoperability issues had been available during the original test on board a ship.

Given the results of the current test, there are additional products interoperable with Products A and F, which were selected as meeting ILO requirements in the original test report. Specifically the list of products that are interoperable with sufficient performance averaged over native and interoperable usage now includes Products A, C and F.

All future vendors attempting to produce biometric systems for use with ILO SID applications should be provided with the guidance document from Annex B of this document.

Amendments should eventually be undertaken to ISO 19794-2 and reflected in ILO SID-0002 to reflect the following guidance to resolve ambiguities:

1. Matchers claiming the ability to use the standardized format should be able to match even if some or all of the minutiae in a data interchange format are marked as Type '00' or "Other".
2. Minutiae encoders claiming the ability to produce minutiae records in accordance with the standard should not only encode the minutiae angle using the quantization level specified in the standard, but should attempt to calculate the angle as closely as possible to that quantization level.
3. Minutiae truncation, where necessary, should use the method suggested in Annex B of this standard.

References

a. ILO Seafarers' Identity Documents Biometric Testing Campaign Report - Part 1

b. Seafarers' Identity Documents Convention (Revised), 2003 (Convention No. 185)

c. ILO SID-0002 Finger Minutiae-Based Biometric Profile for the Seafarers' Identity Documents

d. ISO/IEC CD 19794-2 – Biometric Data Interchange Formats – Part 2: Finger Minutiae Data (ISO/IEC JTC 1 SC37 N 340, dated 2003-10-07)

Acknowledgments

The authors of this report would like to gratefully acknowledge the National Biometric Security Project (NBSP) for financially supporting the offline test and preparation of the report, ISO/IEC JTC 1 SC 37 for providing a group of experts to review this report, and Crystal Cruises, along with the crew and officers of the Crystal Harmony for graciously agreeing to participate in the original ILO Biometric Testing Campaign.

Glossary

An attempt has been made to harmonize the definitions and terms used in this report with common industry practice and with the various reference standards listed above. Specific relevant terms are defined below for the reader's convenience.

Attempt
Submission of one (or a sequence of) biometric samples to the system.
NOTE 1: An attempt results in a matching score (or scores), or possibly a 'failure to acquire'.
NOTE 2: Many biometric systems collect and process a sequence of samples in a single attempt, for example:
(a) collecting samples over some fixed period, and scoring the best matching sample;
(b) collecting samples until either a match is obtained or the system times out;
(c) collecting samples until one of sufficient quality is obtained, or the system times out;
(d) collecting a second sample when the score from the first sample is very close to the decision threshold.

Biometric
Pertaining to the field of biometrics, used as an adjective.

Biometric information record (BIR)
A data structure containing a BDB, information identifying the BDB format, and possibly further information such as whether the BDB is digitally signed or encrypted.

Biometric verification/biometrically verify
Validate that a biometric sample matches the previously-stored processed biometric sample associated with the subject's claimed identity by comparing the templates, generating a score, and comparing the score with the threshold.

Biometric enrollment
The process of collecting one or more biometric samples from a subject and the subsequent preparation and storage of one or more processed biometric samples and associated data representing that subject's identity.

Detection error tradeoff curve (DET curve)
Curve which plots error rates on both axes (false positives on the x-axis and false negatives on the y-axis), which is usually plotted using logarithmic axes.
NOTE 1: DET curves can be used to plot matching error rates (false non-match rate against false match rate), decision error rates (false reject rate against false accept rate), open-set identification error rates (false negative identification rate against false positive identification rate). In this last case, the curves will depend on (a) the number of users enrolled in the database, and (b) the number of identifiers returned (per attempt) by the identification system.

Failure-to-acquire rate (FTA)
Expected proportion of attempts for which the system is unable to capture or locate an image or signal of sufficient quality.

False Accept Rate (FAR)
Expected proportion of verification transactions with wrongful claims of identity (in a positive identification system) or non-identity (in a negative identification system) that are incorrectly confirmed by the biometric system.

False Reject Rate (FRR)
Expected probability that a biometric verification system will incorrectly not match a person’s biometric sample to his/her biometric reference data (in a positive identification system) or incorrectly match a person’s biometric sample to someone else’s biometric reference data (in a negative identification system); thereby denying truthful claims of identity.
NOTE 1: See clause 4.3.2 for further details.
NOTE 2: In this document, biometric reference data are finger minutiae-templates as described in ILO SID-0002, Annex II.

Genuine attempt
Single good-faith attempt by a user to match their own biometric reference data.
NOTE: In this document, biometric reference data are finger minutiae-templates as described in ILO SID-0002, Annex II.

Template
User’s stored reference measure based on features extracted from enrollment samples.
NOTE: In this document, the template consists of minutiae data and other information for a single fingerprint as described in Annex B of ILO SID-0002 as either the 1st Fingerprint or 2nd Fingerprint Sections within the BIR.

Annex A – Technical Contribution from NBSP (Draft)

Summary of Interoperability Issues Surrounding ILO SID-0002 and ISO 19794-2

Preliminary Draft for Review

John Campbell
Bion Biometrics
January 5, 2004

Introduction

During the recent ILO Biometric Technology Tests held during September and October of 2004, seven different biometric products, each consisting of a fingerprint sensor together with a minutiae extraction and matching algorithm were tested. All of the products were required to enroll fingerprints using the template format specified in ILO SID-0002, which is essential the normal sized card format specified in ISO 19794-2, with the number of minutiae truncated to 52. This format does not allow any space for core, delta or ridge count information or for proprietary extended information; simply five bytes per minutiae for up to 52 minutiae per finger. All tests were performed with live seafarers to more accurately simulate an operational environment. Thus, each verification attempt was made using a seafarer presenting their finger to a specific product which was attempting to match them to the standardized enrollment template created either by the same or a different product. The details of ILO SID-002 are at http://www.ilo.org/public/english/dialogue/sector/papers/maritime/sid0002.pdf and the details of the test performed and the results obtained are currently available at http://www.ilo.org/public/english/dialogue/sector/papers/maritime/sid-test-report1.pdf

Without describing the results in detail (as they are well described in the test report mentioned above), it was obvious that six out of the seven products worked well provided they were making verification attempts on standardized templates they had created themselves, but performance (as measured by FRR at a fixed FAR of 1% or 2%) was highly variable (from 0% to 73%) when attempting to verify seafarers against templates that had been created by other vendors. This suggested that either the standard was defined too loosely (or was simply confusing and difficult to implement correctly), or the vendors had made significant implementation errors in the products provided for testing.

The National Biometrics Security Project in the US was sufficiently concerned about the possible implications of this for other implementations of ISO 19794-2 and the corresponding US standard ANSI INCITS 378-2004, that they agreed to support an effort to liaise with NIST, SC 37 experts and the original product vendors to try and investigate this issue further. After lengthy discussions with vendors, many of which are confidential because they involved details about algorithm implementations, the list of issues shown below has been developed. It is expected that the vendors who previously had problems will now produce new versions of their software and a new test will occur using the previously collected database of fingerprint images. Although this test will not involve live seafarer interaction, it will use the same images gathered from the previous interaction with the sensors and will thus be a reasonable simulation of how the products would have performed if the refined algorithms had been available in the original scenario test. It is acknowledged, of course, that human factors such as frustration with a particular product if it produced multiple failures to match on genuine verification attempts, may have influenced the images captured, and no changes to the algorithms can now rectify that, but this is the closest test that the authors know how to achieve to a perfect repetition of the original test.

The issues can be classified into three categories. First, there are those issues that require modifications to the base standard or strict conformance requirements in a companion conformance testing standard. Fortunately, an initial analysis of which vendors performed well together indicates that many of these are not likely to have been the main factors that impacted interoperability performance. Second, there are those issues that result from subtleties in the standard that may be quite adequately defined, but that many vendors either ignored or did not understand. Guidance documentation, perhaps in a corresponding conformance testing standard or in an application profile, needs to be provided to indicate the importance of these issues. Once again, however, some of these issues appear not to be the largest sources of error, since vendors who interpreted them differently were still able to interoperate. Finally, there are those areas where one or two vendors appear to have made mistakes during implementation. These were apparently described properly and with enough emphasis in the standard, since most vendors interpreted them correctly, but a set of caveats will be listed in this document. It can be provided to other vendors seeking to implement the standard so that they will be aware of specific areas to pay extra attention to during their implementation.

Issues Requiring Standard Modification or Strict Conformance Requirements

Minutiae Type

ISO 19794-2 allows for minutiae to be labeled as ridge endings, ridge bifurcations or “other”. There is no clear definition of the “other” type and although it is suggested in the standard that “other” type minutiae be counted as matching minutiae if they overlap with a minutiae of any of the three types, it is left to the vendor to determine how this affects the score. In fact, it appears that the minutiae extraction and truncation algorithms in different products may produce significantly different quantities of each type of minutiae. The seven products tested varied from all “other” (one product) to no “other” (six products) and from an almost even split between endpoints and bifurcations to a preference for one minutiae type of up to 70%, when all templates across all fingers enrolled from the test group of seafarers were included. Some vendors have admitted that their scores will be lower when matching against “other” minutiae or against minutiae of the wrong type (e.g. bifurcation instead of endpoint). The immediately obvious solution would be to amend the standard to eliminate or severely restrict the use of "other" minutiae. It seems, however, that due to the large differences between percentages of endpoints and bifurcations among different vendors, it may be more appropriate to make this recommendation:

1) It should be a conformance requirement to ISO 19794-2 that any matcher that uses templates conforming to the standard be able to function accurately even without knowledge of minutiae type in the enrollment template.

In this case, minutiae type would be an extra piece of information that vendors could use when they felt it was advantageous, perhaps because the source of the originating template was known from the CBEFF header, but which would not be required in order to achieve interoperable matching.

Minutiae Truncation

Document 19794-2 and previous drafts have discussed various strategies for the truncation of minutia to allow for small data format size. The current text included in ILO SID-0002 was "*If the number of minutiae exceeds the maximum number processible by a card, truncation is necessary. The truncation is a 2 step process. At first, finger minutiae of poor quality are eliminated. If still too many minutiae are there, then truncation shall be made by peeling off minutiae from the convex hull of the minutiae set and before sorting into the order required by the card.*"

Since minutiae quality is not defined in 19794-2, every vendor has a different method of assigning minutiae quality and deciding which minutiae to truncate first. Several vendors also expressed that they felt the meaning of "peeling off minutiae from the convex hull of the minutiae set" was not clear. In the end, every vendor decided to use their own proprietary truncation method. This is definitely a problem for applications where truncation will be required.

Using the assumption that vendors will always remove some minutiae for having extremely low quality, it seems reasonable to accept that minutiae quality algorithms will be proprietary (since they must depend on the minutiae extraction techniques used) and delete that portion from the truncation section. A mathematically simpler truncation strategy could then be adopted. This leads to the following recommendation.

2) Either by amending 19794-2 or by adding guidance to biometric profiles or specific procurement documents, replace the current text on truncation from the last paragraph of 8.3.1 in SC 37 N0464 with "*If the number of minutiae after all normal minutiae extraction steps exceeds the maximum number processible by a card, truncation is necessary. The truncation should remove minutiae in order of largest distance from the centroid of the current minutiae set, with the centroid being recalculated after each minutiae is removed. This truncation operation should occur prior to sorting the minutiae into any specific order required by the card.*"

Given that several product pairs did achieve reasonable interoperability, it seems that this may not be a major issue in achieving interoperability. The test to be performed shortly will hopefully shed more light on this, as vendors are not being asked to change their truncation algorithms.

Minutia Direction

It is not clear from the standard exactly how minutiae angle is defined. The standard gives two figures and a description based on a skeletonized image, but this may not be useful to those vendors who don't normally skeletonize to a single pixel ridge skeleton. There are two alternatives here. The first is to try and exactly define in the standard a specific means of computing an angle for each minutiae point. This, however, may not be equally applicable to all minutiae extraction algorithms and may bias the standard in favour of certain technologies. The second is to leave the definition as it is, but provide guidance that vendors try to allow enough flexibility in their matching algorithms to accommodate slightly different interpretations of minutiae angle. Since at least two vendors have revealed that they used quantized angles, but these were still able to interoperate well with some other vendors, it is likely that minor differences in minutiae direction may not be an issue. Thus, until the next test is complete, no recommendation is required on this issue. If angle is identified to cause problems after the next test, then an amendment to the angle definition in the standard may be required.

Issues Requiring Further Guidance Beyond Existing Standard

Minutiae Position

The ISO 19794-2 standard has six different formats. Even the normal sized card format has two different variants, one in which the endpoint minutiae are defined as ridge skeleton endpoints and one in which they are defined as valley skeleton bifurcation points. Since there may be several pixels difference in the position of a minutiae depending on which of these definitions is used, it is possible for this to lead to interoperability problems if vendors don't implement in the manner prescribed by the particular variant of the standard being used. Nominally, the particular format used is defined by a CBEFF format type, but several vendors expressed confusion with this. Thus, it would be beneficial to somewhere have a detailed table explaining exactly the differences between the different format types and referencing those parts of the standard that should be reviewed to understand these differences. This would be appropriate for inclusion in a biometric application profile or even in a procurement document for a specific deployment. The recommendation is:

3) Provide guidance in application specific profiles or procurement documents to indicate the specific 19794-2 format types, with reference to the specific paragraphs in the standard defining their characteristics, that vendors are expected to support. This would include definition of endpoint minutiae, inclusion of extended data, etc.

There is one positive note here. The two vendors who were most compatible in the ILO test used different definitions of endpoint minutiae and so it seems that this is not a major factor affecting interoperability.

Endianness

ISO 19794-2 is specifically designated as being Big-Endian. It is important for vendors to be reminded of that, since their programmers may have already developed a favoured Endianness when implementing native templates. Even more important in the short term is the fact that BioAPI (as defined in ANSI INCITS 358-2002) is defined to be Little-Endian. Thus, if minutiae templates are being wrapped inside a BioAPI CBEFF header, part of the data record will be Little-Endian and part will be Big-Endian. This caused problems for almost every vendor. Thus the recommendation is:

4) For any biometric application profile or procurement specification that attempts to reference minutiae templates within the context of BioAPI, guidance should be provided to the vendors to indicate that the BioAPI CBEFF header is Little-Endian while the minutiae record is Big-Endian.

Fortunately, the current international version of BioAPI (ISO 19784-1) has changed to Big-Endian and when this becomes widely deployed, this issue should become irrelevant.

Angle Quantization

The angle in the normal sized card format is encoded into 8 bits. Each bit is supposed to be quantized to represent an angle increment of 2*pi/256. Two of the seven vendors, however, specifically indicated that they used a different quantization internally prior to encoding the angles. These two vendors had significant interoperability problems. Although it is not known if the angle quantization was specifically the problem in this case, it is probably worth including a note on this in guidance for the use of the standard. Thus the recommendation is:

5) Vendors should be reminded in biometric application profiles or procurement documents to pay close attention to the quantization requirements specified for the relevant format types in the standard (e.g. 2*pi/256 for minutiae angle in the card normal format) and to bear that final quantization in mind as they extract values and write them to the minutiae record.

Issues Requiring Vendor Care During Implementation

Minutiae Ordering

Unless a specific value is provided in a Biometric Information Template header, it is assumed that there is no specific order to the minutiae. Some vendors, however, do order the minutiae and may expect minutiae to be in a particular order. This definitely affects interoperability. Thus it is very important to ensure that the version of a matching algorithm used to implement this standard does not require minutiae in a particular order.

Coordinate System

Although the coordinate system for minutiae positions and angles is well defined in ISO 19794-2, several vendors seemed to have problems implementing it correctly the first time. All vendors would be recommended to double check their coordinate systems by using some type of external display tool, since coordinate transforms within the algorithm will often still allow matching with the same algorithm. It is frequently only during interoperability testing that coordinate transform issues become significant. This may be a particular issue because some vendors may have implemented coordinate transforms to deal with particular sensor imperfections, and it is important these are implemented properly to produce minutiae positions in the absolute coordinate system specified in the standard.

Translation Invariance

Since standardized templates may be produced using many different sensors, there will certainly be the potential for a wide variety in sensor sizes. Although the coordinate system for the minutiae position is specified in absolute coordinates, it is tied to the physical corner of the sensor. Thus a fingerprint centered in a very large sensor will have a large translation from the same fingerprint centered in a small sensor. Although most algorithms accommodate translation, the tolerances may need to be loosened to support this type of interoperability.

Sensor Resolution

Even with a minimum sensor resolution of 500 dpi, as was specified in ILO SID-0002, there is no rule that prohibits some templates from being produced on sensors with resolution greater than 500 dpi. The image size in the X and Y direction, however, is defined in the standard as being the size of the original image in pixels. All of the vendors interpreted this as meaning the original size of the raw image produced by the sensor. When displaying the image for review, or computing which minutiae are close to the edge of the image (which can be used to determine if a minutiae is a so called "edge minutiae"), it is important to multiply the image size by the correct X and Y resolution factors in order to convert it to the same coordinate system as the minutiae. Failure to do so can cause some minutiae to appear to be close to or even beyond the edge of the image.

These factors seem to have been the cause of particular problems in the implementations for this test and hopefully being forewarned of them will assist other vendors who plan to develop implementations of ILO SID-0002, ISO 19794-2 or ANSI INCITS 378-2004.

Annex B – Technical Contribution from NBSP

Summary of Interoperability Issues Surrounding ILO SID-0002 and ISO 19794-2

Technical Contribution from National Biometric Security Project (NBSP)

Author: John Campbell, Bion Biometrics

Revision 1.0
January 31, 2004

Introduction

During the recent ILO Biometric Technology Tests held during September and October of 2004, seven different biometric products, each consisting of a fingerprint sensor together with a minutiae extraction and matching algorithm were tested. All of the products were required to enroll fingerprints using the template format specified in ILO SID-0002, which is essential the normal sized card format specified in ISO 19794-2, with the number of minutiae truncated to 52. This format does not allow any space for core, delta or ridge count information or for proprietary extended information; simply five bytes per minutiae for up to 52 minutiae per finger. All tests were performed with live seafarers to more accurately simulate an operational environment. Thus, each verification attempt was made using a seafarer presenting their finger to a specific product which was attempting to match them to the standardized enrollment template created either by the same or a different product. The details of ILO SID-002 are at http://www.ilo.org/public/english/dialogue/sector/papers/maritime/sid0002.pdf and the details of the test performed and the results obtained are currently available at http://www.ilo.org/public/english/dialogue/sector/papers/maritime/sid-test-report1.pdf

Without describing the results in detail (as they are well described in the test report mentioned above), it was obvious that six out of the seven products worked well provided they were making verification attempts on standardized templates they had created themselves, but performance (as measured by FRR at a fixed FAR of 1% or 2%) was highly variable (from 0% to 73%) when attempting to verify seafarers against templates that had been created by other vendors. This suggested that either the standard was defined too loosely (or was simply confusing and difficult to implement correctly), or the vendors had made significant implementation errors in the products provided for testing.

The National Biometrics Security Project in the US was sufficiently concerned about the possible implications of this for other implementations of ISO 19794-2 and the corresponding US standard ANSI INCITS 378-2004, that they agreed to support an effort to liaise with NIST, SC 37 experts and the original product vendors to try and investigate this issue further. After lengthy discussions with vendors, many of which are confidential because they involved details about algorithm implementations, the list of issues shown below has been developed. It is expected that the vendors who previously had problems will now produce new versions of their software and a new test will occur using the previously collected database of fingerprint images. Although this test will not involve live seafarer interaction, it will use the same images gathered from the previous interaction with the sensors and will thus be a reasonable simulation of how the products would have performed if the refined algorithms had been available in the original scenario test. It is acknowledged, of course, that human factors such as frustration with a particular product if it produced multiple failures to match on genuine verification attempts, may have influenced the images captured, and no changes to the algorithms can now rectify that, but this is the closest test that the authors know how to achieve to a perfect repetition of the original test.

The issues can be classified into three categories. First, there are those issues that require modifications to the base standard or strict conformance requirements in a companion conformance testing standard. Fortunately, an initial analysis of which vendors performed well together indicates that many of these are not likely to have been the main factors that impacted interoperability performance. Second, there are those issues that result from subtleties in the standard that may be quite adequately defined, but that many vendors either ignored or did not understand. Guidance documentation, perhaps in a corresponding conformance testing standard or in an application profile, needs to be provided to indicate the importance of these issues. Once again, however, some of these issues appear not to be the largest sources of error, since vendors who interpreted them differently were still able to interoperate. Finally, there are those areas where one or two vendors appear to have made mistakes during implementation. These were apparently described properly and with enough emphasis in the standard, since most vendors interpreted them correctly, but a set of caveats will be listed in this document. It can be provided to other vendors seeking to implement the standard so that they will be aware of specific areas to pay extra attention to during their implementation.

Issues Requiring Standard Modification or Strict Conformance Requirements

Minutiae Type

ISO 19794-2 allows for minutiae to be labeled as ridge endings, ridge bifurcations or “other”. There is no clear definition of the “other” type and although it is suggested in the standard that “other” type minutiae be counted as matching minutiae if they overlap with a minutiae of any of the three types, it is left to the vendor to determine how this affects the score. In fact, it appears that the minutiae extraction and truncation algorithms in different products may produce significantly different quantities of each type of minutiae. The seven products tested varied from all “other” (one product) to no “other” (six products) and from an almost even split between endpoints and bifurcations to a preference for one minutiae type of up to 70%, when all templates across all fingers enrolled from the test group of seafarers were included. Some vendors have admitted that their scores will be lower when matching against "other" minutiae or against minutiae of the wrong type (e.g. bifurcation instead of endpoint). The immediately obvious solution would be to amend the standard to eliminate or severely restrict the use of "other" minutiae. It seems, however, that due to the large differences between percentages of endpoints and bifurcations among different vendors, it may be more appropriate to make this recommendation:

6) It should be a conformance requirement to ISO 19794-2 that any matcher that uses templates conforming to the standard be able to function accurately even without knowledge of minutiae type in the enrollment template.

In this case, minutiae type would be an extra piece of information that vendors could use when they felt it was advantageous, perhaps because the source of the originating template was known from the CBEFF header, but which would not be required in order to achieve interoperable matching.

Minutiae Truncation

Document 19794-2 and previous drafts have discussed various strategies for the truncation of minutia to allow for small data format size. The current text included in ILO SID-0002 was "*If the number of minutiae exceeds the maximum number processible by a card, truncation is necessary. The truncation is a 2 step process. At first, finger minutiae of poor quality are eliminated. If still too many minutiae are there, then truncation shall be made by peeling off minutiae from the convex hull of the minutiae set and before sorting into the order required by the card.*"

Since minutiae quality is not defined in 19794-2, every vendor has a different method of assigning minutiae quality and deciding which minutiae to truncate first. Several vendors also expressed that they felt the meaning of "peeling off minutiae from the convex hull of the minutiae set" was not clear. In the end, every vendor decided to use their own proprietary truncation method. This is definitely a problem for applications where truncation will be required.

Using the assumption that vendors will always remove some minutiae for having extremely low quality, it seems reasonable to accept that minutiae quality algorithms will be proprietary (since they must depend on the minutiae extraction techniques used) and delete that portion from the truncation section. A mathematically simpler truncation strategy could then be adopted. This leads to the following recommendation.

7) Either by amending 19794-2 or by adding guidance to biometric profiles or specific procurement documents, replace the current text on truncation from the last paragraph of 8.3.1 in SC 37 N0464 with "*If the number of minutiae after all normal minutiae extraction steps exceeds the maximum number processible by a card, truncation is necessary. The truncation should remove minutiae in order of largest distance from the centroid of the minutiae set. If multiple minutiae have the same distance from the centroid and not all of them need too be truncated in order to reach the maximum number of minutiae, then truncate the group of*

same distance minutiae in the order that they are found in the image. First truncate same distance minutiae with the lowest x-coordinate, and if they have the same x-coordinate then truncate starting with the lowest y-coordinate. This truncation operation should occur prior to sorting the minutiae into any specific order required by the card."

Given that several product pairs did achieve reasonable interoperability, it seems that this may not be a major issue in achieving interoperability. The test to be performed shortly will hopefully shed more light on this, as vendors are not being asked to change their truncation algorithms.

Minutia Direction

It is not clear from the standard exactly how minutiae angle is defined. The standard gives two figures and a description based on a skeletonized image, but this may not be useful to those vendors who don't normally skeletonize to a single pixel ridge skeleton. There are two alternatives here. The first is to try and exactly define in the standard a specific means of computing an angle for each minutiae point. This, however, may not be equally applicable to all minutiae extraction algorithms and may bias the standard in favour of certain technologies. The second is to leave the definition as it is, but provide guidance that vendors try to allow enough flexibility in their matching algorithms to accommodate slightly different interpretations of minutiae angle. Since at least two vendors have revealed that they used quantized angles, but these were still able to interoperate well with some other vendors, it is likely that minor differences in minutiae direction may not be an issue. Thus, until the next test is complete, no recommendation is required on this issue. If angle is identified to cause problems after the next test, then an amendment to the angle definition in the standard may be required.

Issues Requiring Further Guidance Beyond Existing Standard

Minutiae Position

The ISO 19794-2 standard has six different formats. Even the normal sized card format has two different variants, one in which the endpoint minutiae are defined as ridge skeleton endpoints and one in which they are defined as valley skeleton bifurcation points. Since there may be several pixels difference in the position of a minutiae depending on which of these definitions is used, it is possible for this to lead to interoperability problems if vendors don't implement in the manner prescribed by the particular variant of the standard being used. Nominally, the particular format used is defined by a CBEFF format type, but several vendors expressed confusion with this. Thus, it would be beneficial to somewhere have a detailed table explaining exactly the differences between the different format types and referencing those parts of the standard that should be reviewed to understand these differences. This would be appropriate for inclusion in a biometric application profile or even in a procurement document for a specific deployment. The recommendation is:

8) Provide guidance in application specific profiles or procurement documents to indicate the specific 19794-2 format types, with reference to the specific paragraphs in the standard defining their characteristics, that vendors are expected to support. This would include definition of endpoint minutiae, inclusion of extended data, etc.

There is one positive note here. The two vendors who were most compatible in the ILO test used different definitions of endpoint minutiae and so it seems that this is not a major factor affecting interoperability.

Endianness

ISO 19794-2 is specifically designated as being Big-Endian. It is important for vendors to be reminded of that, since their programmers may have already developed a favoured Endianness when implementing native templates. Even more important in the short term is the fact that BioAPI (as defined in ANSI INCITS 358-2002) is defined to be Little-Endian. Thus, if minutiae templates are being wrapped inside a BioAPI CBEFF header, part of the data record will be Little-Endian and part will be Big-Endian. This caused problems for almost every vendor. Thus the recommendation is:

9) For any biometric application profile or procurement specification that attempts to reference minutiae templates within the context of BioAPI, guidance should be provided to the vendors to indicate that the BioAPI CBEFF header is Little-Endian while the minutiae record is Big-Endian.

Fortunately, the current international version of BioAPI (ISO 19784-1) has changed to Big-Endian and when this becomes widely deployed, this issue should become irrelevant.

Angle Quantization

The angle in the normal sized card format is encoded into 8 bits. Each bit is supposed to be quantized to represent an angle increment of 2*pi/256. Two of the seven vendors, however, specifically indicated that they used a different quantization internally prior to encoding the angles. These two vendors had significant interoperability problems. Although it is not known if the angle quantization was specifically the problem in this case, it is probably worth including a note on this in guidance for the use of the standard. Thus the recommendation is:

10) Vendors should be reminded in biometric application profiles or procurement documents to pay close attention to the quantization requirements specified for the relevant format types in the standard (e.g. 2*pi/256 for minutiae angle in the card normal format) and to bear that final quantization in mind as they extract values and write them to the minutiae record.

Issues Requiring Vendor Care During Implementation

Minutiae Ordering

Unless a specific value is provided in a Biometric Information Template header, it is assumed that there is no specific order to the minutiae. Some vendors, however, do order the minutiae and may expect minutiae to be in a particular order. This definitely affects interoperability. Thus it is very important to ensure that the version of a matching algorithm used to implement this standard does not require minutiae in a particular order.

Coordinate System

Although the coordinate system for minutiae positions and angles is well defined in ISO 19794-2, several vendors seemed to have problems implementing it correctly the first time. All vendors would be recommended to double check their coordinate systems by using some type of external display tool, since coordinate transforms within the algorithm will often still allow matching with the same algorithm. It is frequently only during interoperability testing that coordinate transform issues become significant. This may be a particular issue because some vendors may have implemented coordinate transforms to deal with particular sensor imperfections, and it is important these are implemented properly to produce minutiae positions in the absolute coordinate system specified in the standard.

Translation Invariance

Since standardized templates may be produced using many different sensors, there will certainly be the potential for a wide variety in sensor sizes. Although the coordinate system for the minutiae position is specified in absolute coordinates, it is tied to the physical corner of the sensor. Thus a fingerprint centered in a very large sensor will have a large translation from the same fingerprint centered in a small sensor. Although most algorithms accommodate translation, the tolerances may need to be loosened to support this type of interoperability.

Sensor Resolution

Even with a minimum sensor resolution of 500 dpi, as was specified in ILO SID-0002, there is no rule that prohibits some templates from being produced on sensors with resolution greater than 500 dpi. The image size in the X and Y direction, however, is defined in the standard as being the size of the original image in pixels. All of the vendors interpreted this as meaning the original size of the raw image produced by the sensor. When displaying the image for review, or computing which minutiae are close to the edge of the image (which can be used to determine if a minutiae is a so called "edge minutiae"), it is important to multiply the image size by the correct X and Y resolution factors in order to convert it to the same coordinate system as the minutiae. Failure to do so can cause some minutiae to appear to be close to or even beyond the edge of the image.

These factors seem to have been the cause of particular problems in the implementations for this test and hopefully being forewarned of them will assist other vendors who plan to develop implementations of ILO SID-0002, ISO 19794-2 or ANSI INCITS 378-2004.

Annex C – Interoperability Tables

Table 10. Baseline Test: Dual-Finger FRR at 2% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	47.2%	0.9%	34.5%	2.0%	42.8%
B	1.7%	1.7%	2.6%	5.0%	3.1%	5.1%
C	0.0%	3.1%	0.3%	2.9%	64.2%	4.8%
E	0.9%	3.5%	2.5%	1.8%	2.8%	4.7%
F	0.3%	25.1%	0.6%	23.1%	0.3%	33.5%
G	1.7%	6.3%	2.7%	10.8%	1.7%	5.9%

Table 11. Baseline Test #2: Dual-Finger FRR at 2% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	47.3%	0.6%	34.5%	2.0%	42.8%
B	1.4%	1.4%	1.4%	3.9%	2.8%	4.9%
C	0.0%	2.7%	0.6%	2.6%	64.2%	5.4%
E	0.9%	3.1%	2.5%	2.0%	2.8%	4.8%
F	0.3%	25.1%	0.6%	23.1%	0.3%	33.5%
G	1.7%	8.1%	3.0%	8.3%	2.0%	100%*

Table 12. Enhanced Test: Dual-Finger FRR at 2% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	2.3%	0.6%	2.6%	0.3%	2.0%
B	2.0%	1.2%	2.9%	2.6%	1.4%	2.6%
C	0.0%	2.9%	0.3%	2.6%	0.3%	2.0%
E	0.6%	1.2%	1.9%	1.6%	0.6%	0.6%
F	0.3%	2.8%	0.6%	2.5%	0.0%	3.4%
G	0.6%	2.0%	1.8%	2.0%	0.9%	1.7%

Table 13. Original Test: Dual-Finger FRR at 2% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	1.9%	49.1%	7.0%	38.3%	1.4%	40.7%
B	2.7%	1.5%	7.8%	4.2%	4.5%	2.7%
C	0.0%	5.9%	2.2%	2.4%	53.0%	2.0%
E	2.7%	2.4%	5.4%	3.2%	2.4%	2.0%
F	3.0%	38.3%	5.4%	29.1%	0.3%	25.9%
G	2.7%	4.7%	36.1%	8.1%	14.6%	1.5%

Table 14. Difference (Original - Baseline): Dual-Finger FRR at 2% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	3.5%	1.9%	10.4%	3.8%	-0.5%	-2.1%
B	2.3%	-0.2%	6.8%	-0.8%	1.4%	-2.5%
C	0.0%	2.8%	3.6%	-0.4%	-11.1%	-2.8%
E	3.9%	-1.1%	3.4%	1.4%	-0.4%	-2.7%
F	5.7%	13.2%	6.9%	6.0%	0.0%	-7.6%
G	1.4%	-1.6%	34.8%	-2.6%	13.0%	-4.3%

Table 15. Difference (Baseline - Enhanced): Dual-Finger FRR at 2% FAR for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	0.0%	44.9%	0.3%	32.0%	1.7%	40.9%
B	-0.3%	0.5%	-0.3%	2.4%	1.7%	2.5%
C	0.0%	0.3%	0.0%	0.3%	63.9%	2.8%
E	0.3%	2.3%	0.6%	0.3%	2.1%	4.1%
F	0.0%	22.2%	0.0%	20.6%	0.3%	30.1%
G	1.1%	4.3%	0.9%	8.7%	0.8%	4.1%

Table 16. Original Test: Failure To Acquire (FTA) Rates for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	1.6%	0.0%	4.2%	0.0%	0.0%	0.0%
B	1.4%	0.0%	1.6%	0.0%	0.0%	0.0%
C	0.0%	0.0%	1.8%	0.0%	0.0%	0.0%
E	2.1%	0.0%	0.5%	0.0%	0.0%	0.0%
F	3.0%	0.0%	2.1%	0.0%	0.0%	0.0%
G	0.4%	0.0%	1.4%	0.0%	0.0%	0.0%

Table 17. Baseline Test: Genuine Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	348	351	351	352	352	351
B	348	352	351	352	352	352
C	345	349	348	349	346	349
E	321	325	324	325	325	325
F	348	352	351	352	352	352
G	296	300	299	300	300	300

Table 18. Baseline Test: Imposter Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	357	357	355	357	354	357
B	357	357	357	357	357	357
C	351	351	353	351	348	350
E	330	330	330	330	330	330
F	357	357	357	357	357	357
G	303	303	303	303	303	303

Table 19. Baseline Test #2: Genuine Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	348	351	351	352	352	351
B	348	352	351	352	352	352
C	345	349	348	349	346	349
E	321	325	324	325	325	325
F	348	352	351	352	352	352
G	296	300	299	300	300	300

Table 20. Baseline Test #2: Imposter Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	357	357	355	357	354	357
B	357	357	357	357	357	357
C	351	351	353	351	348	350
E	330	330	330	330	330	330
F	357	357	357	357	357	357
G	303	303	303	303	303	303

Table 21. Enhanced Test: Genuine Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	348	351	351	352	352	351
B	342	346	345	346	346	346
C	345	349	348	349	346	349
E	317	322	320	322	319	322
F	348	352	351	352	352	352
G	339	343	342	343	343	343

Table 22. Enhanced Test: Imposter Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	357	357	355	357	354	357
B	351	351	351	351	351	351
C	351	351	353	351	348	350
E	303	303	303	303	303	303
F	357	357	357	357	357	357
G	294	294	294	294	294	294

Table 23. Original Test: Genuine Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	373	188	142	208	208	207
B	222	373	193	188	208	205
C	205	205	370	162	146	205
E	146	192	222	373	166	166
F	167	152	189	194	373	167
G	223	191	145	207	221	373

Table 24. Original Test: Imposter Verifications for All Visits

Enrol Product \ Verify Product	A	B	C	E	F	G
A	363	-	-	-	-	-
B	-	364	-	-	-	-
C	-	-	365	-	-	-
E	-	-	-	364	-	-
F	-	-	-	-	364	-
G	-	-	-	-	-	364

Annex D – Dual Finger DET Curves

Figure 1. Enrol A Verify A

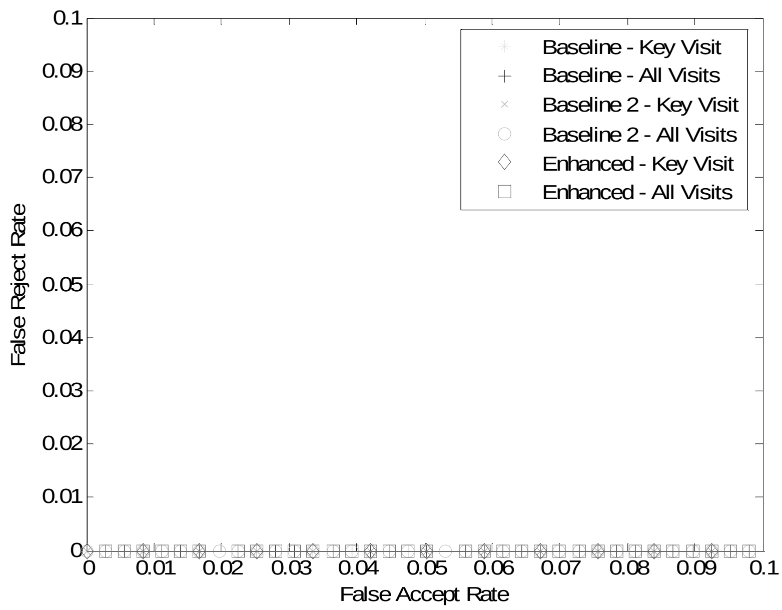


Figure 2. Enrol A Verify B

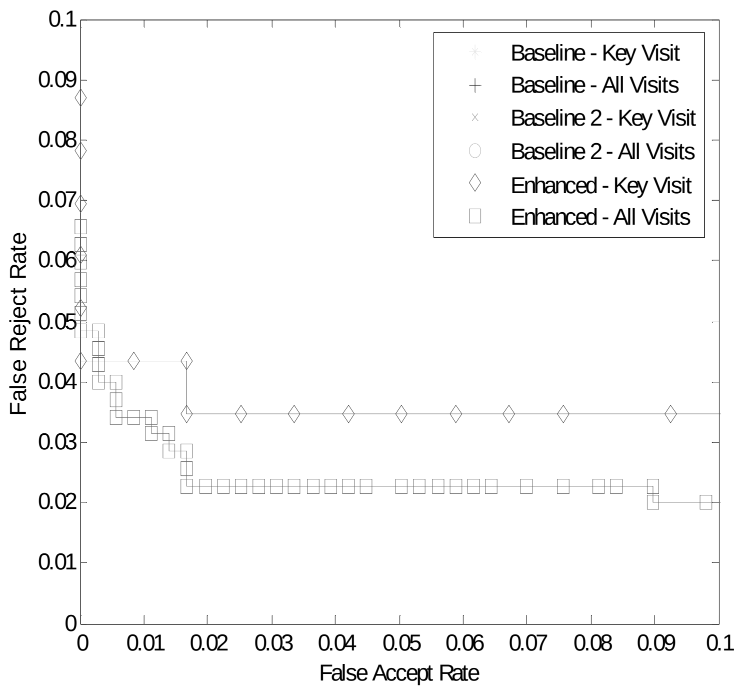


Figure 3. Enrol A Verify C

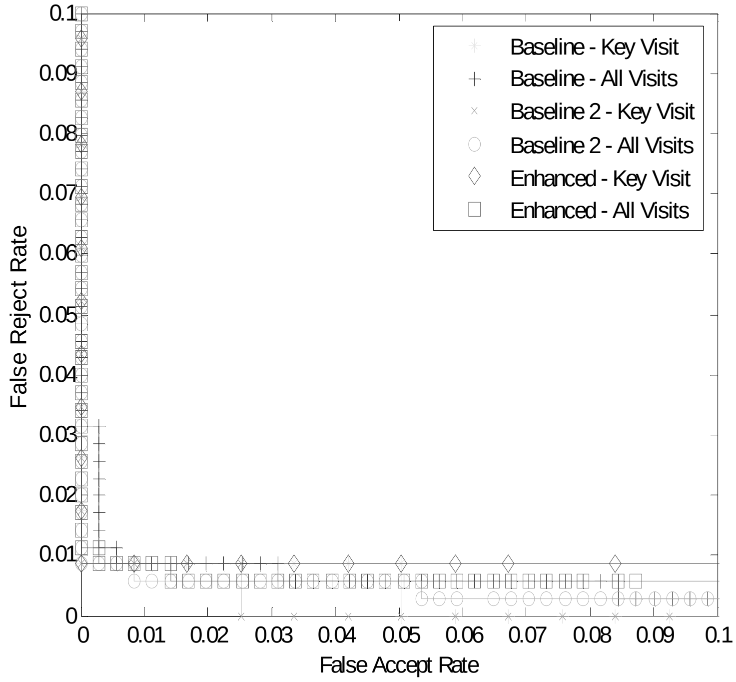


Figure 4. Enrol A Verify E

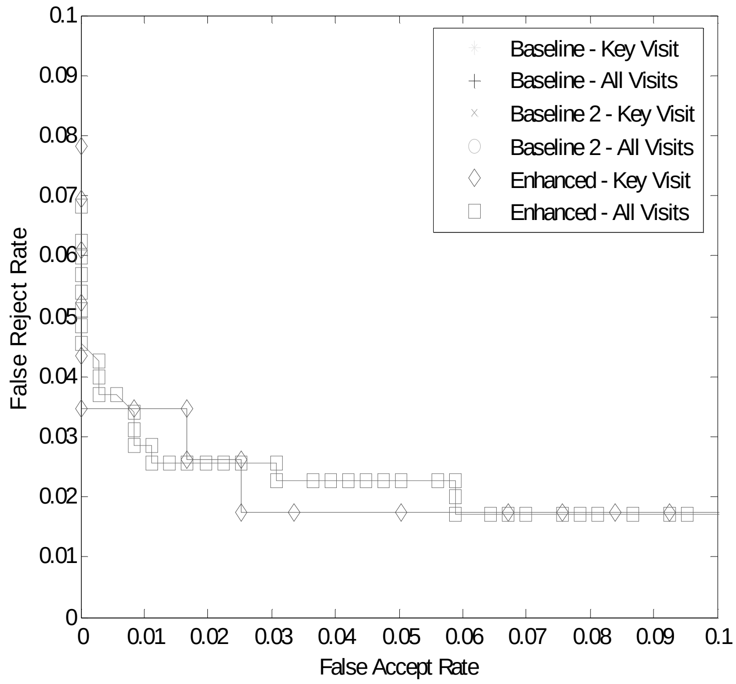


Figure 5. Enrol A Verify F

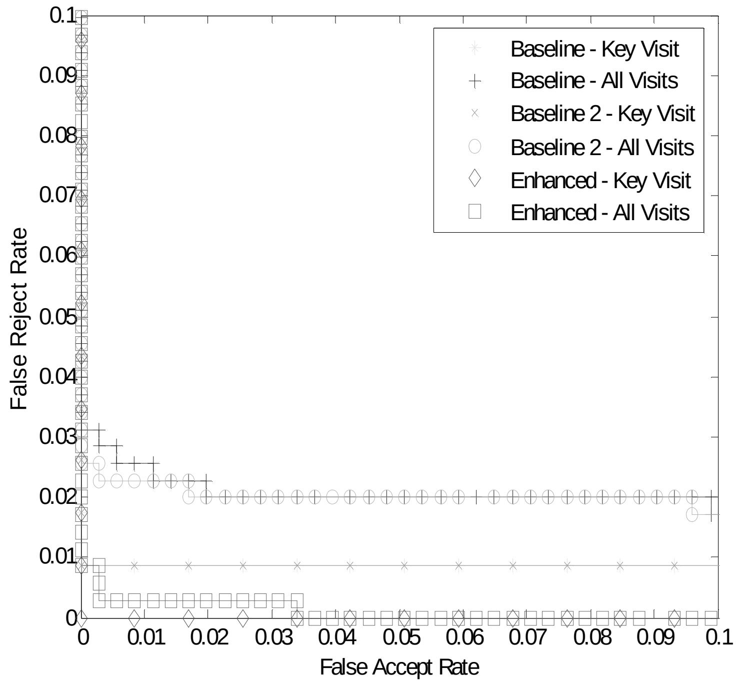


Figure 6. Enrol A Verify G

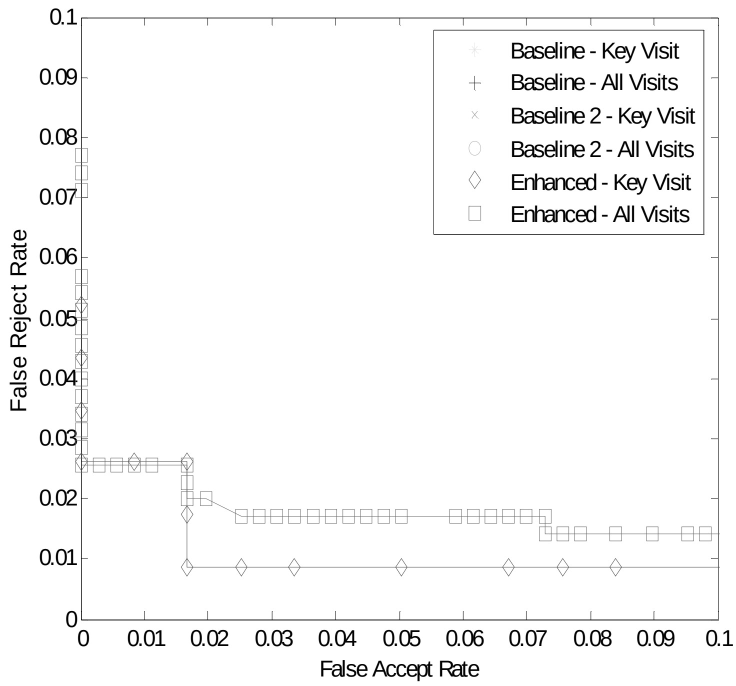


Figure 7. Enrol B Verify A

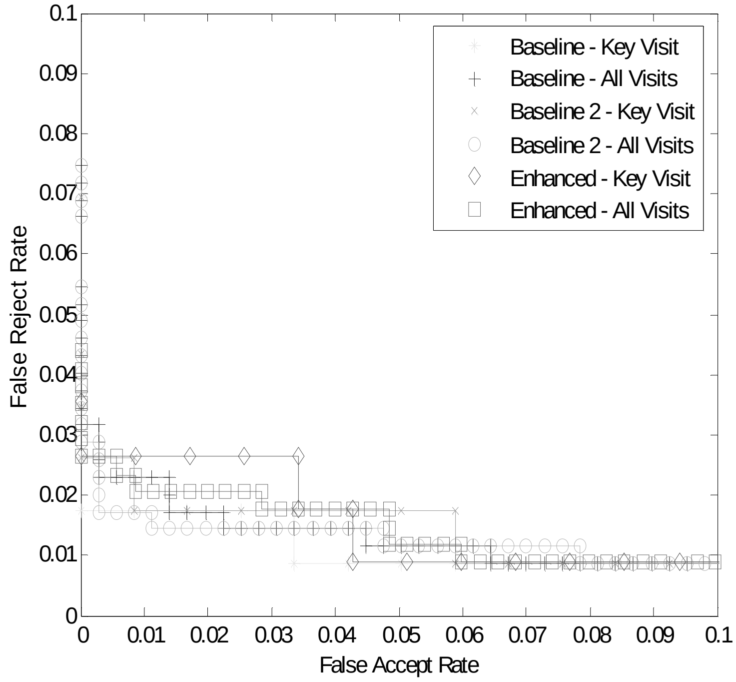


Figure 8. Enrol B Verify B

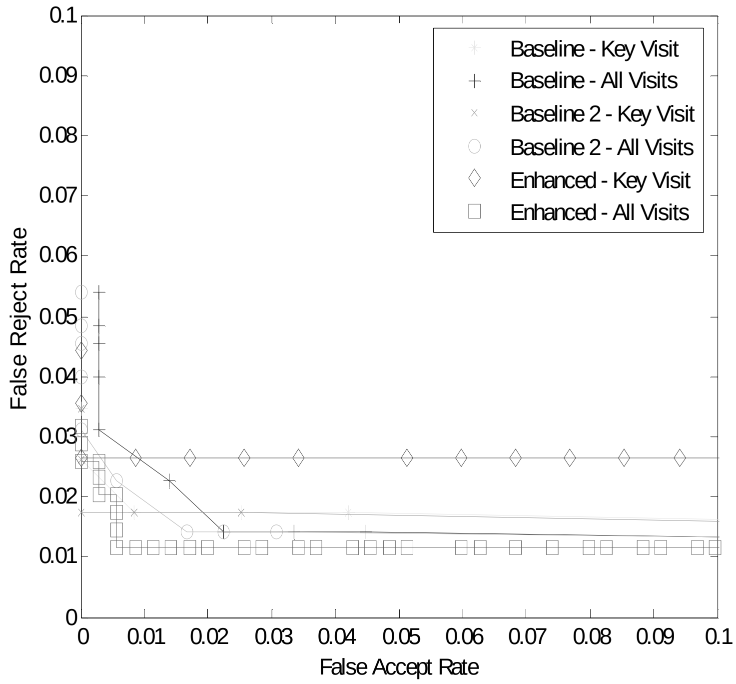


Figure 9. Enrol B Verify C

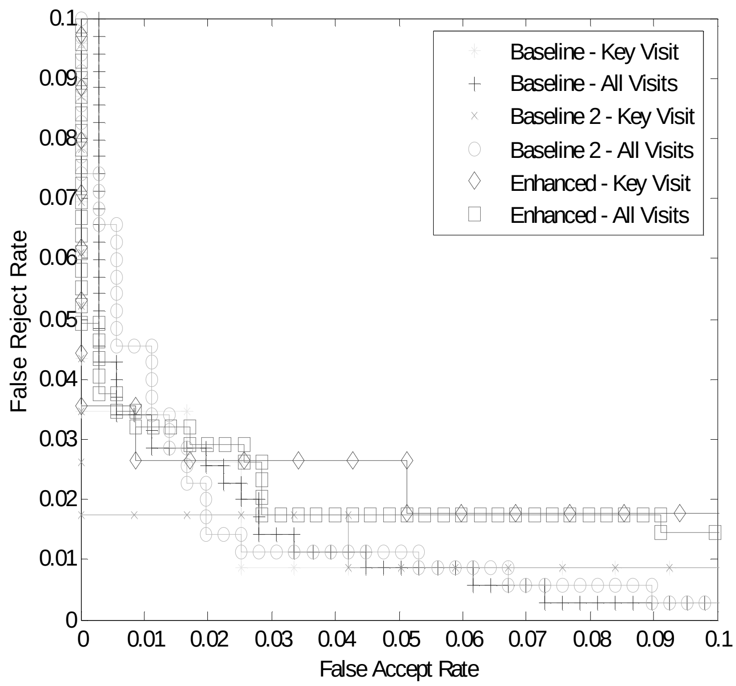


Figure 10. Enrol B Verify E

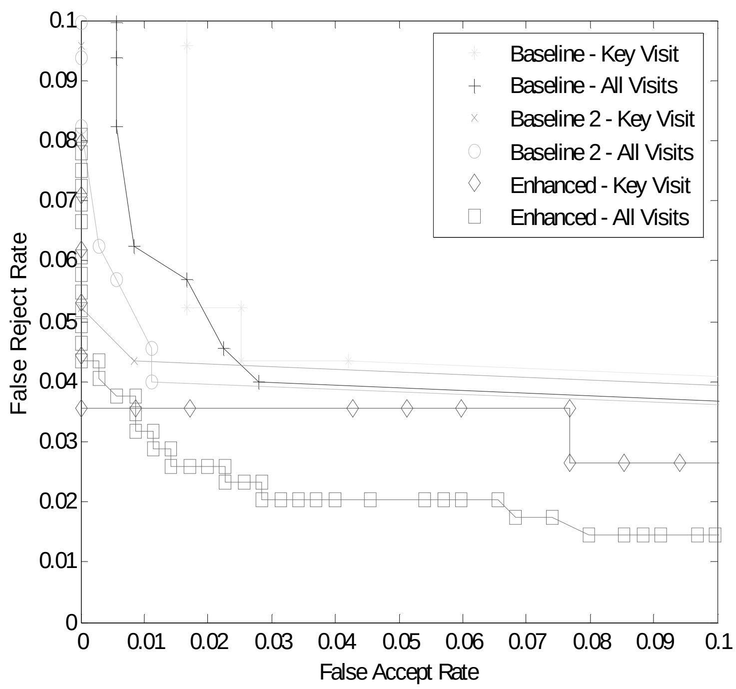


Figure 11. Enrol B Verify F

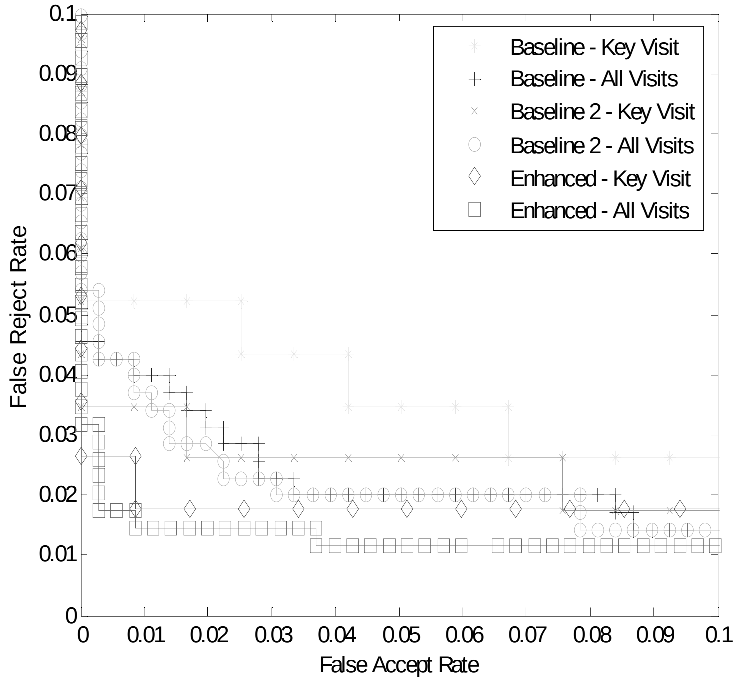


Figure 12. Enrol B Verify G

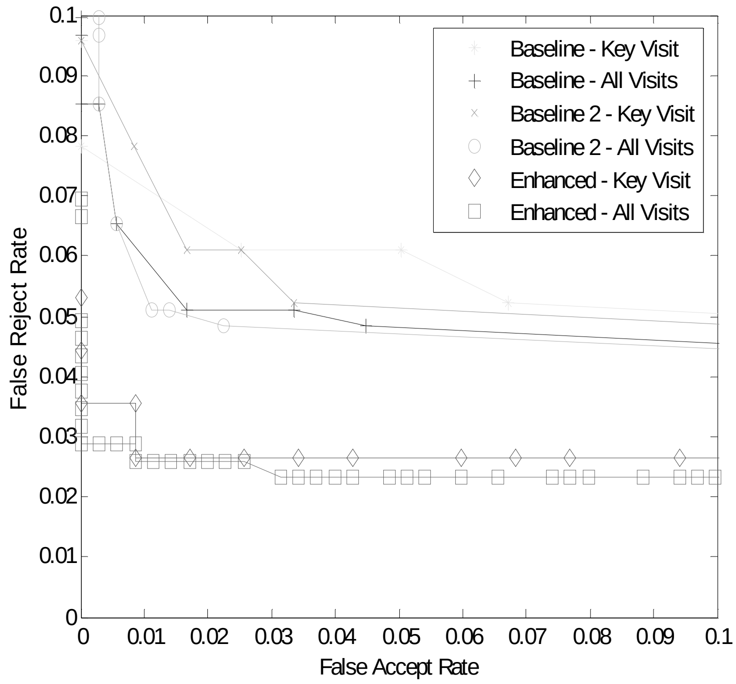


Figure 13. Enrol C Verify A

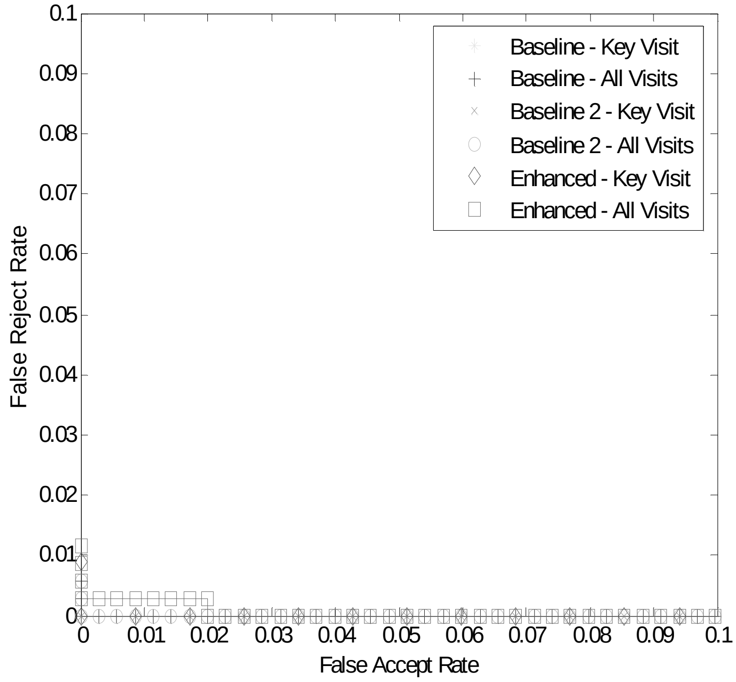


Figure 14. Enrol C Verify B

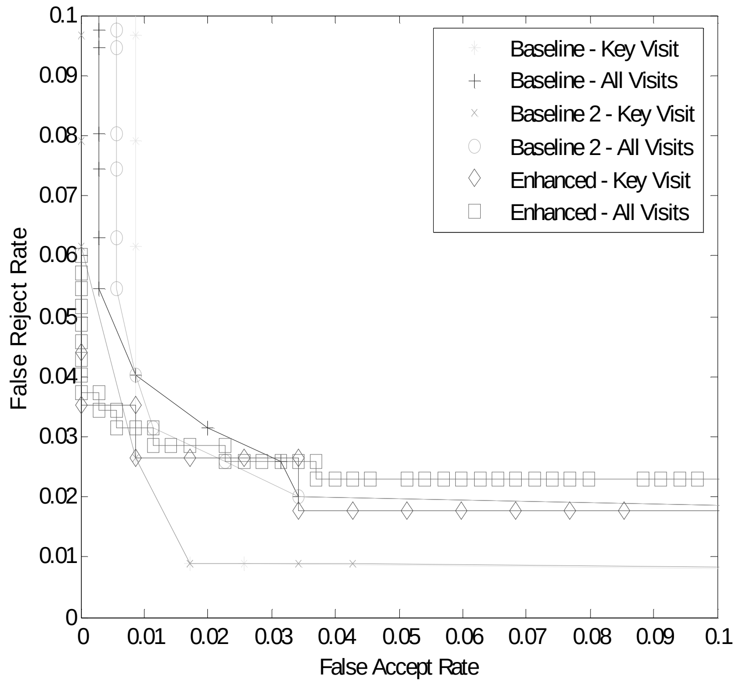


Figure 15. Enrol C Verify C

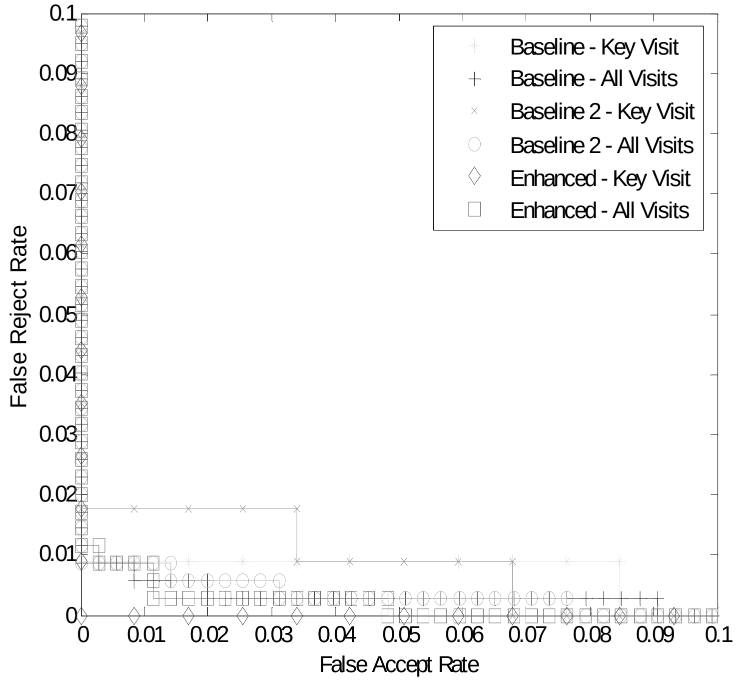


Figure 16. Enrol C Verify E

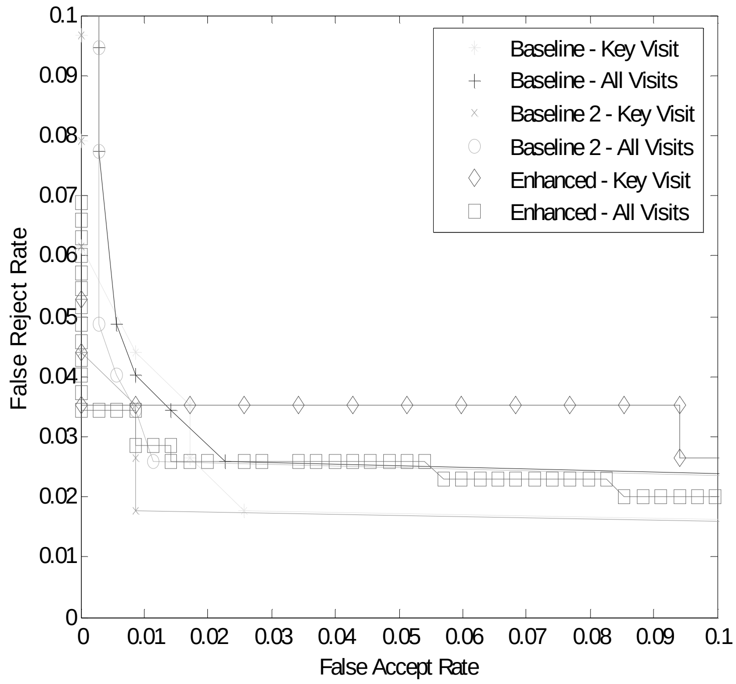


Figure 17. Enrol C Verify F

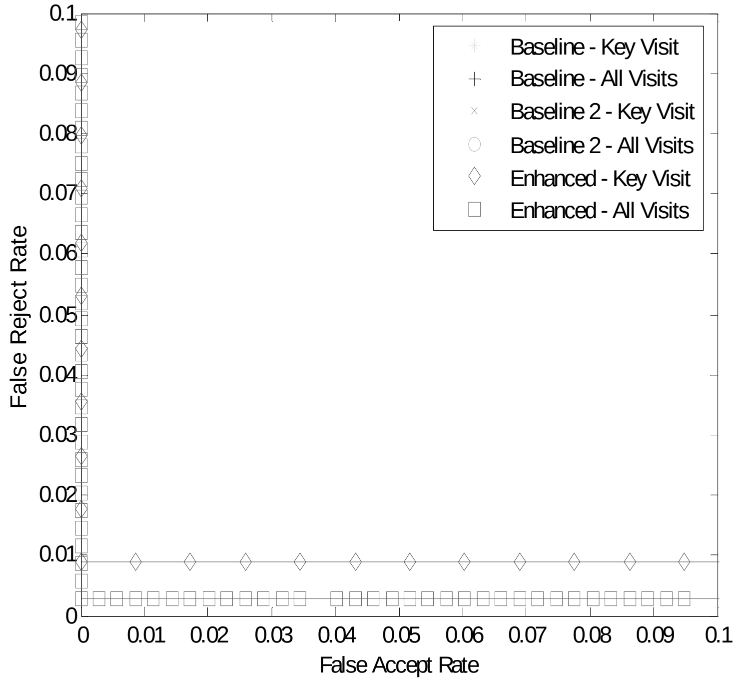


Figure 18. Enrol C Verify G

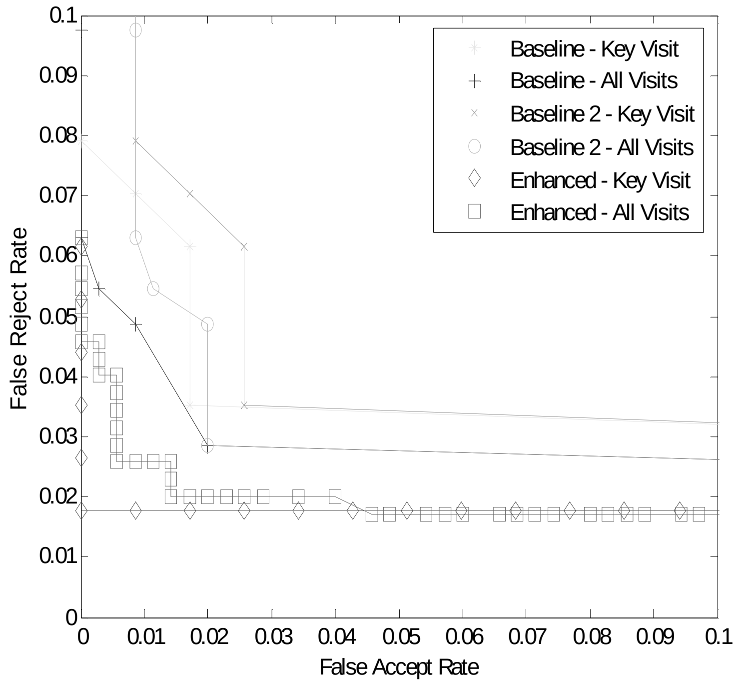


Figure 19. Enrol E Verify A

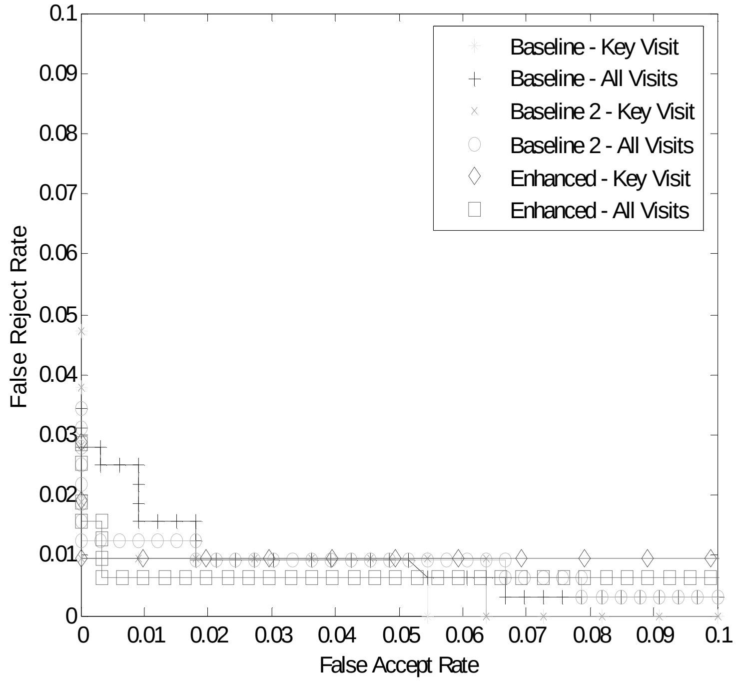


Figure 20. Enrol E Verify B

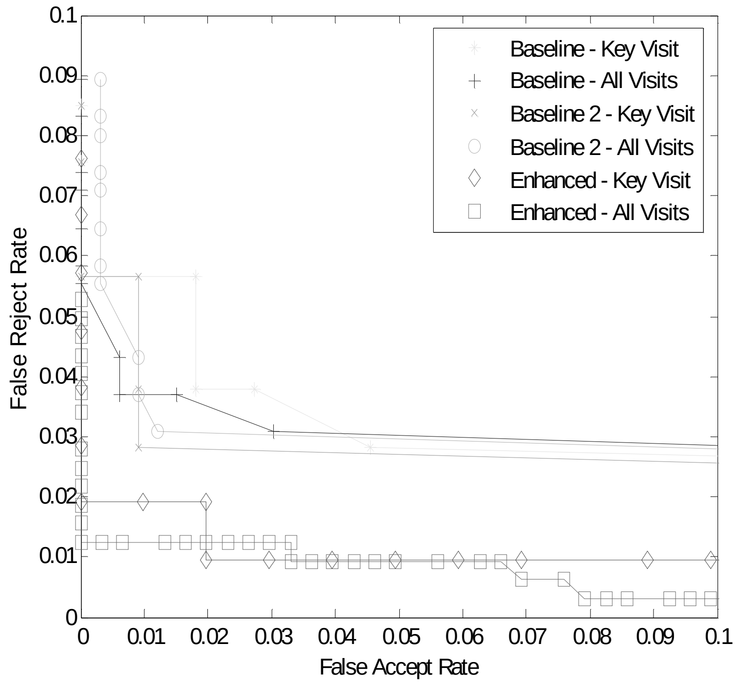


Figure 21. Enrol E Verify C

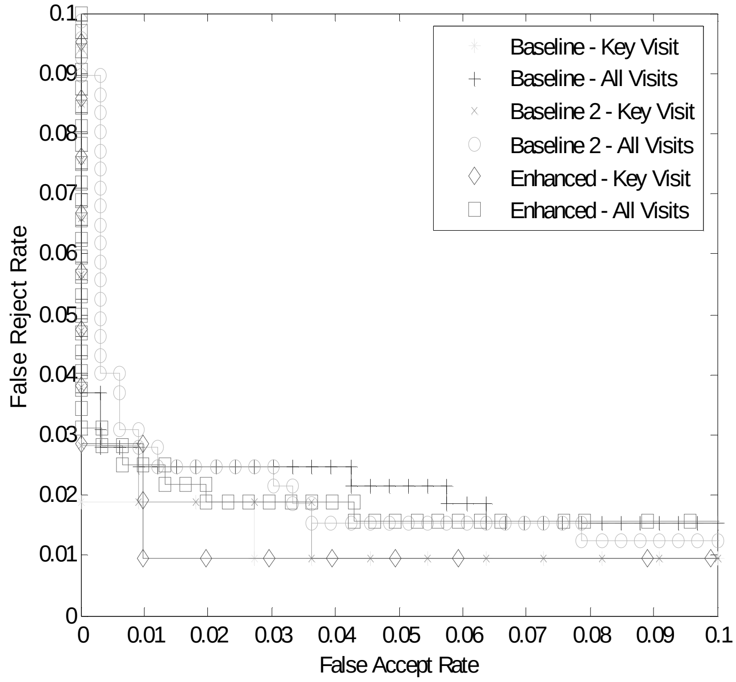


Figure 22. Enrol E Verify E

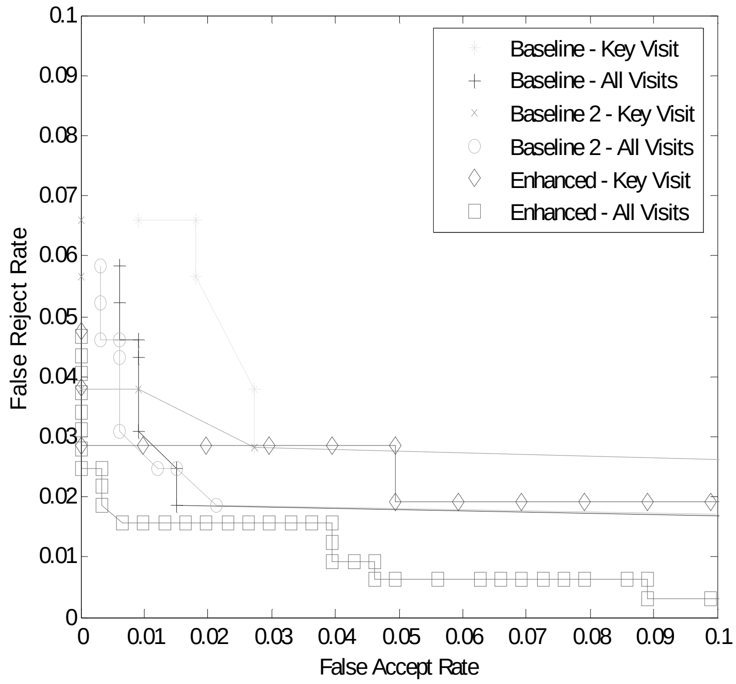


Figure 23. Enrol E Verify F

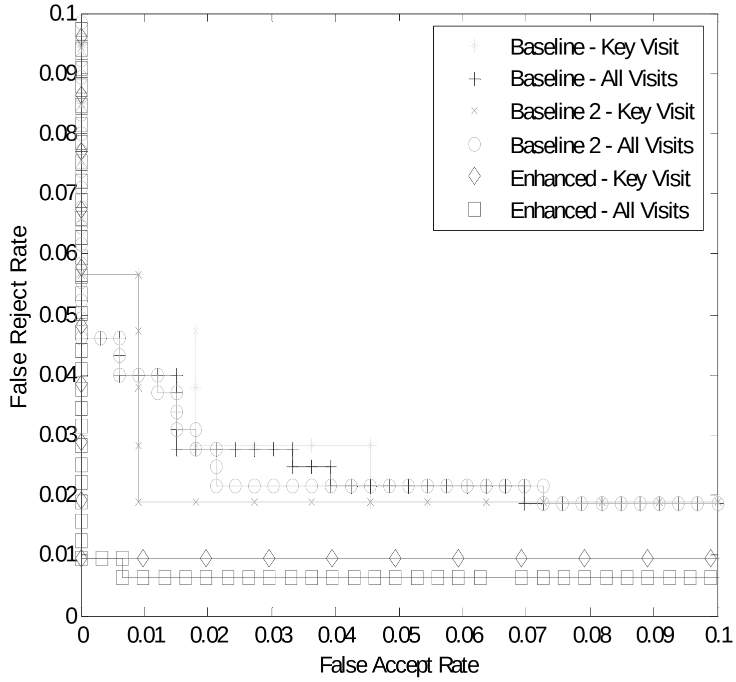


Figure 24. Enrol E Verify G

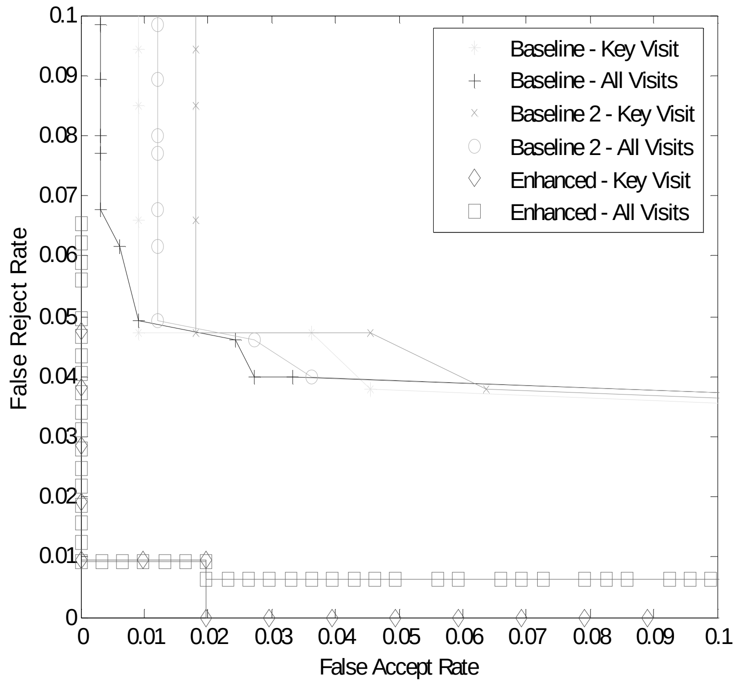


Figure 25. Enrol F Verify A

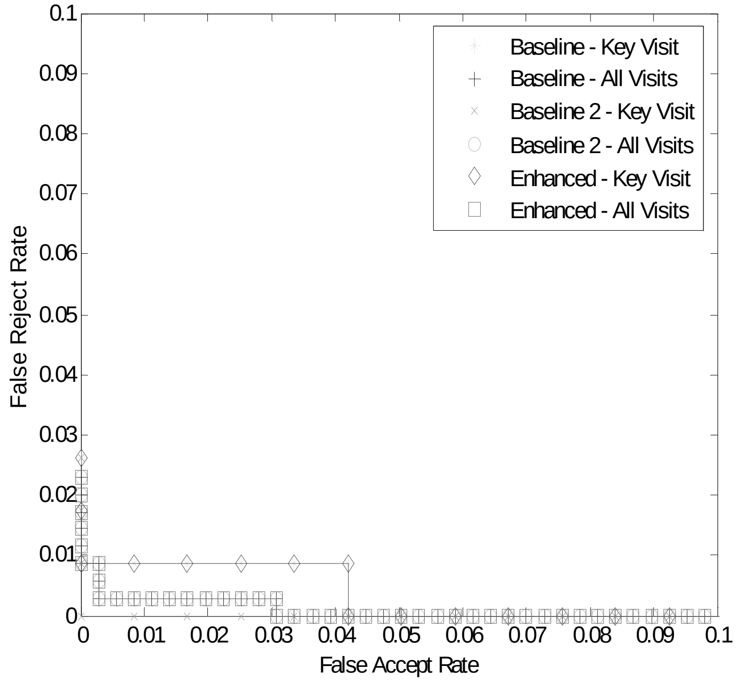


Figure 26. Enrol F Verify B

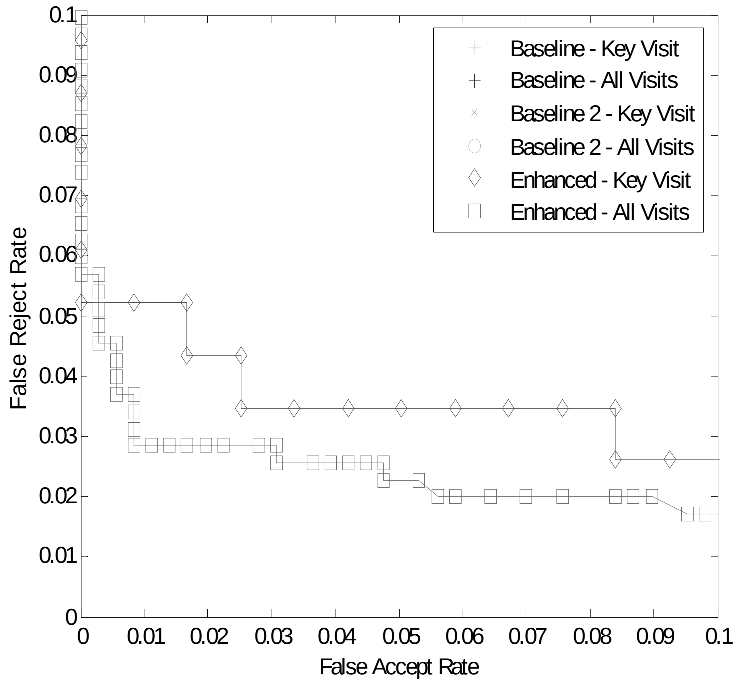


Figure 27. Enrol F Verify C

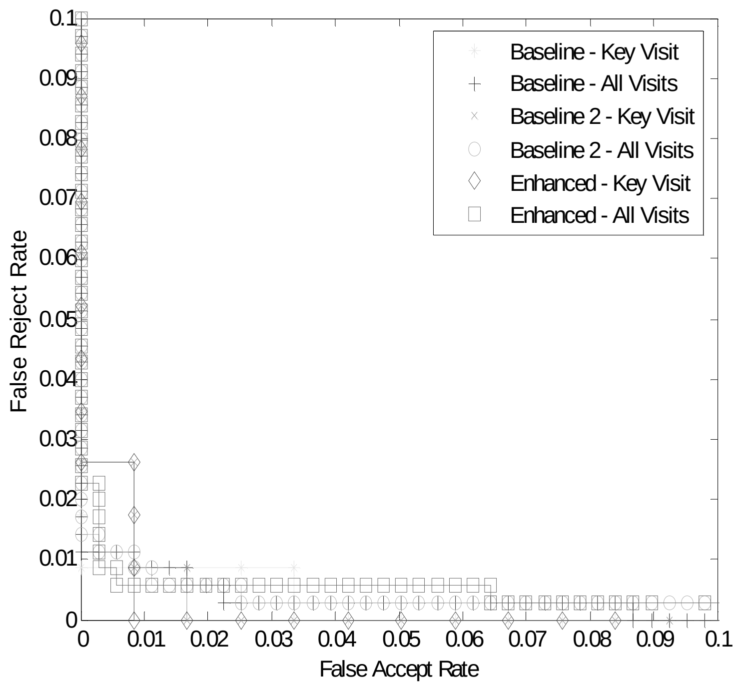


Figure 28. Enrol F Verify E

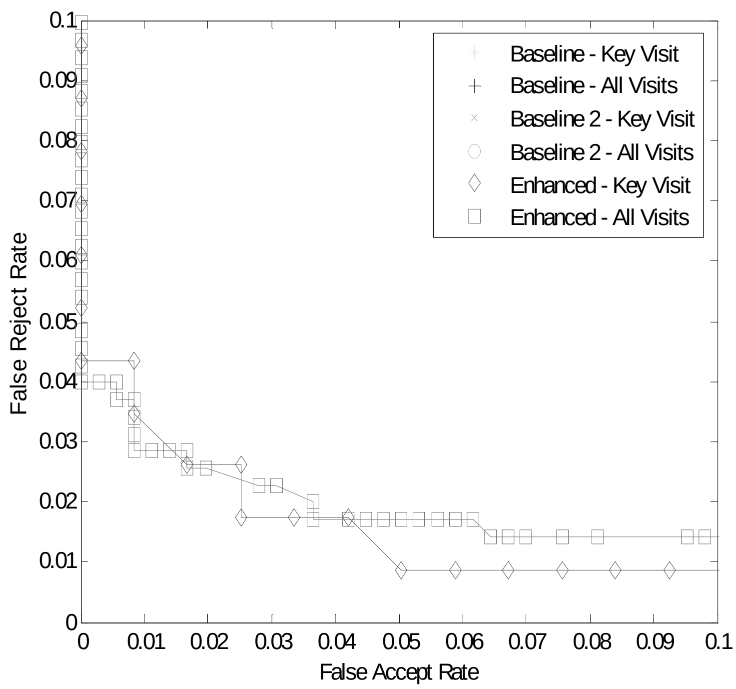


Figure 29. Enrol F Verify F

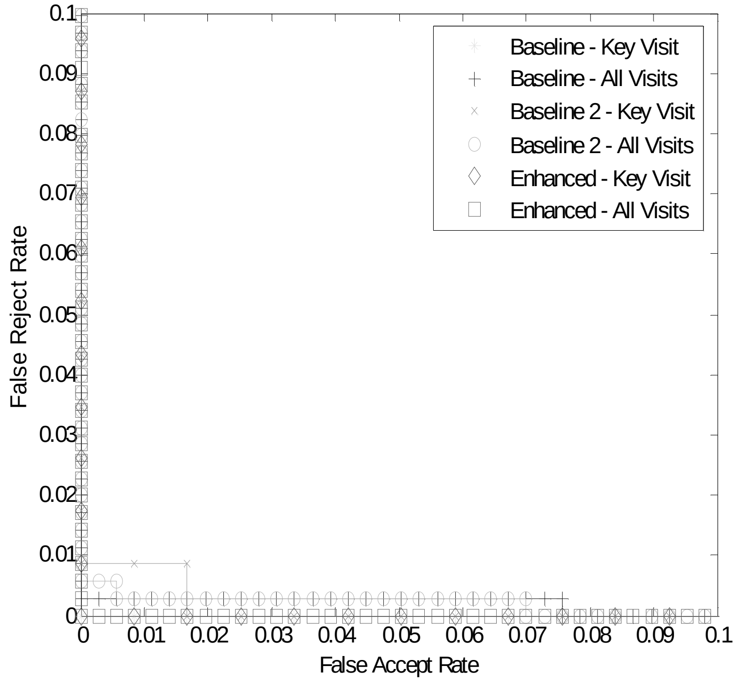


Figure 30. Enrol F Verify G

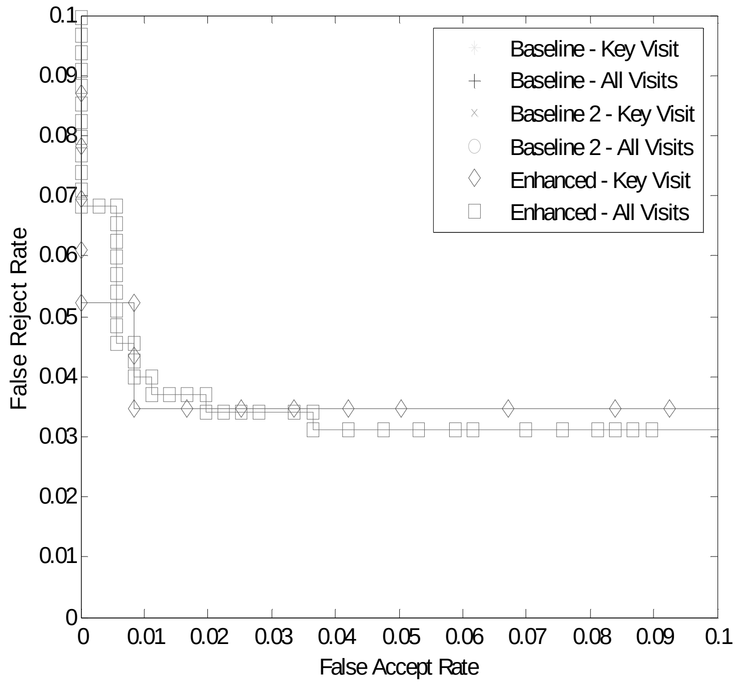


Figure 31. Enrol G Verify A

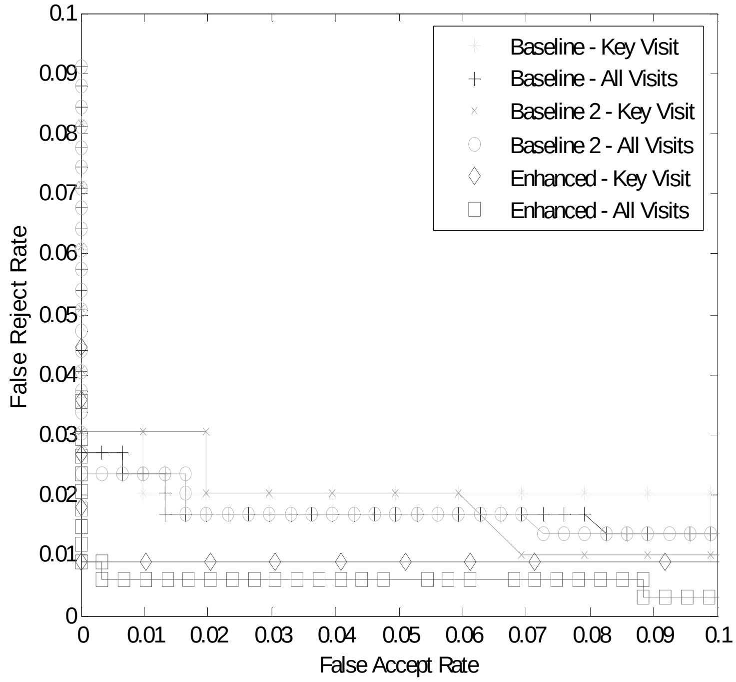


Figure 32. Enrol G Verify B

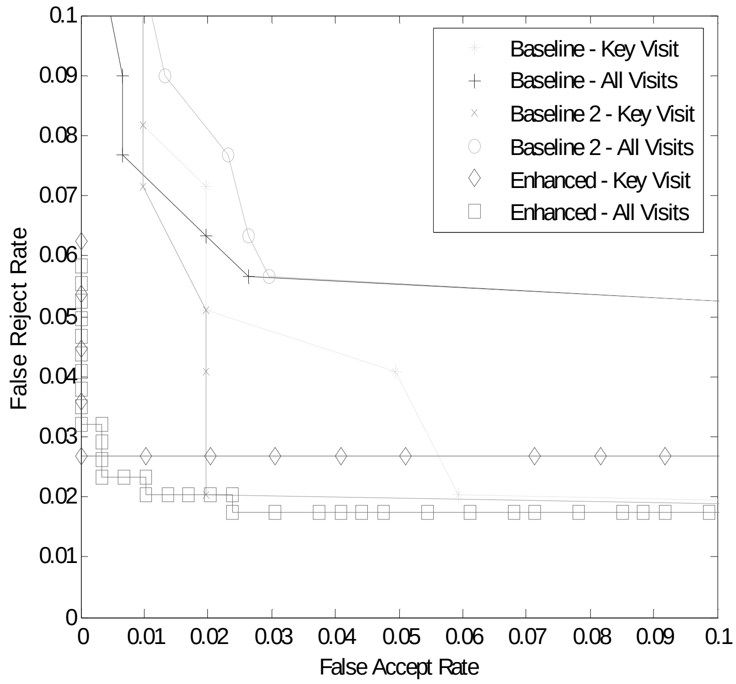


Figure 33. Enrol G Verify C

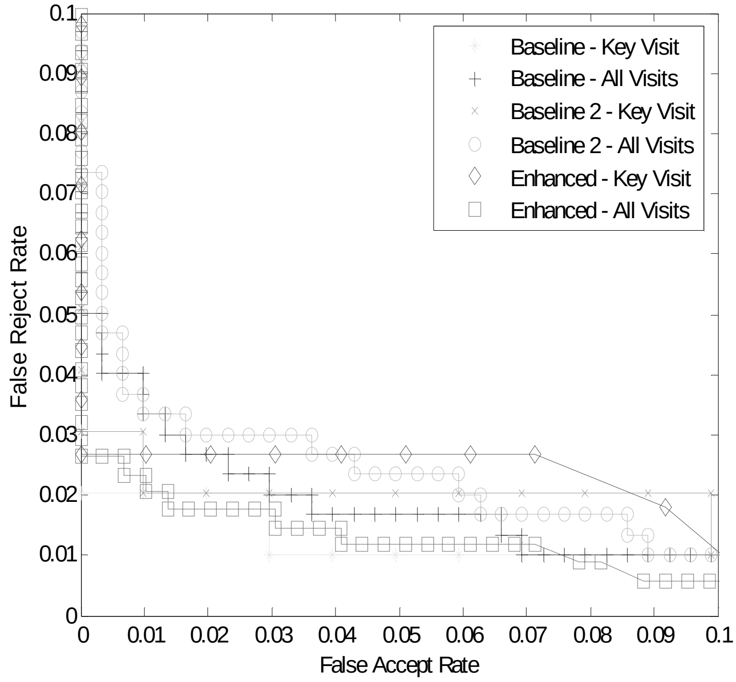


Figure 34. Enrol G Verify E

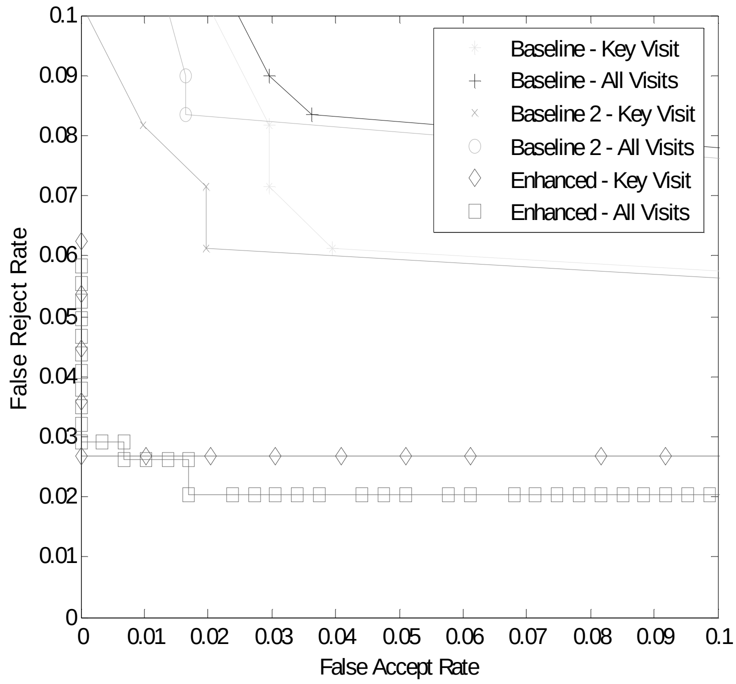


Figure 35. Enrol G Verify F

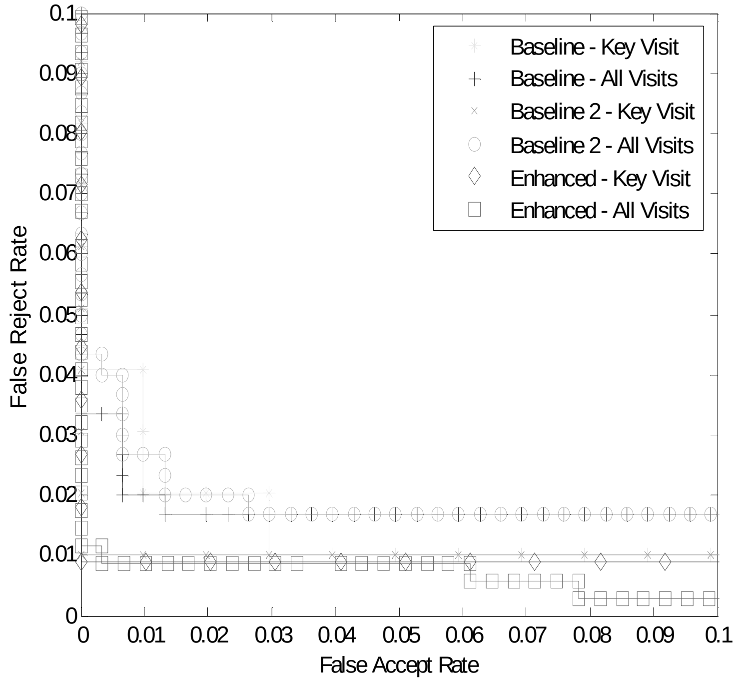
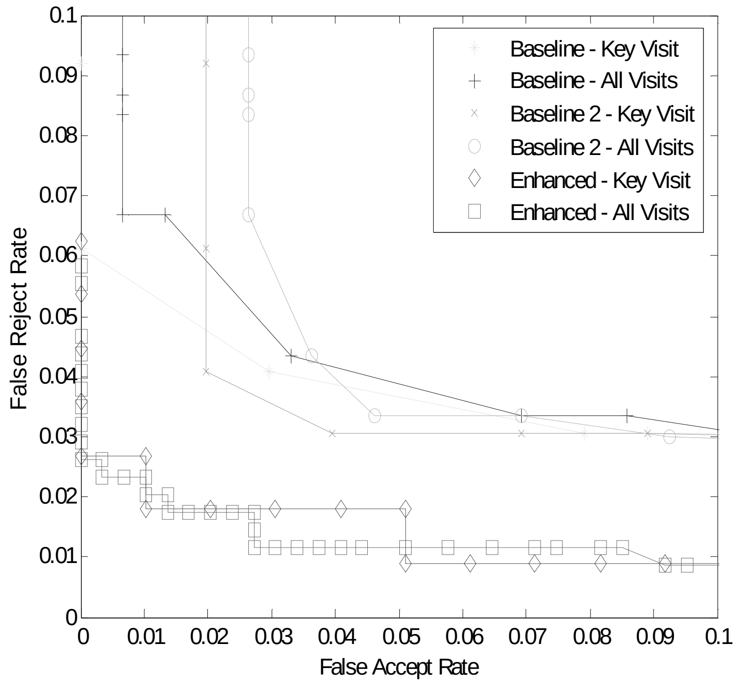


Figure 36. Enrol G Verify G



Annex E – Vendor Software Requirements

Vendor Software Requirements – Option 1 (BioAPI)

BioAPI 1.1 compliant BSP

To participate in off-line testing, vendors must provide a fully BioAPI compliant Biometric Service Provider (BSP) as detailed in section 4.2 of the BioAPI specification, version 1.1. (ANSI/INCITS 358-2002) In addition, vendor BSPs must also supports the following functions:

BioAPI_CreateTemplate

The test application will pass two BMP images (primary finger, followed by secondary) as the opaque data in the input BIR (CapturedBIR). The vendors' BSP must use these images to return an SID-0002 compliant BIR as the enrollment template. This function should NOT output any BIR or raw images to disk.

The BioAPI BIR header of the input BIR (CapturedBIR) will contain the following values:
Length = (16+ length of primary BMP + length of secondary BMP)
BioAPI_BIR_VERSION = 0x01
BioAPI_BIR_DATA_TYPE = 0x01 (BioAPI_BIR_DATA_TYPE_RAW)
BioAPI_BIR_BIOMETRIC_DATA_FORMAT = 0x00000000
BioAPI_QUALITY = 0x00
BioAPI_BIR_PURPOSE = 0x04 (BioAPI_PURPOSE_ENROLL_FOR_VERIFICATION_ONLY)
BioAPI_BIR_AUTH_FACTORS = 0x00000008 (BioAPI_FACTOR_FINGERPRINT)

Note: BSPs should default the new BIR's Finger Position fields to be the right index (2) and left index (7) for the primary and secondary templates respectively. If the images do not correspond to these fingers, the test application will set these fields to the correct values after BioAPI_CreateTemplate returns the new enrollment BIR.

BioAPI_Process

The test application will pass a single BMP image as the opaque data in the input BIR (CapturedBIR). The BSP must use this image to return a processed BIR (of the vendor's choice; this BIR does not have to be SID-0002 compliant) to be matched against an SID-0002 compliant BIR using BioAPI_VerifyMatch. This function should NOT output any BIR or raw images to disk.

The BioAPI BIR header of the input BIR (CapturedBIR) will contain the following values:
Length = (16 + length of raw BMP image)
BioAPI_BIR_VERSION = 0x01
BioAPI_BIR_DATA_TYPE = 0x01 (BioAPI_BIR_DATA_TYPE_RAW)
BioAPI_BIR_BIOMETRIC_DATA_FORMAT = 0x00000000
BioAPI_QUALITY = 0x00
BioAPI_BIR_PURPOSE = 0x01 (BioAPI_PURPOSE_VERIFY)
BioAPI_BIR_AUTH_FACTORS = 0x00000008 (BioAPI_FACTOR_FINGERPRINT)

BioAPI_VerifyMatch

The test application will pass an SID-0002 compliant BIR as the stored enrollment template (StoredTemplate) and the processed BIR (ProcessedBIR) returned from BioAPI_Process. The vendors' BSP should always attempt to match the processed BIR with the primary finger stored in the enrollment template. A match result, and FARAchieved score must be returned.

Raw BMP Images

Vendor BSPs will only be expected to process their own native raw images produced and stored during the 2004 ILO SID Biometric Technology Test. During this test, BSPs should not write BIRs or raw images to disk.

FARAchieved Scores

To properly evaluate a product's performance, a good distribution of scores that follow the BioAPI definition of the BioAPI_FAR type is desired.

MaxFARRequested

Vendors shall clearly indicate in writing, their recommended value of BioAPI_FAR MaxFARRequested to use to define a successful match when a BioAPI_VerifyMatch operation is executed by the test application. If the FAR threshold is set internally, the BSP must accept the default MaxFARRequested value of 1.

System Stability

All vendor BSPs must coexist on a single client PC, so each vendor must mitigate against memory leaks or other issues that could cause instability of the test system.

Unconnected Hardware

No sensors will be connected or used for the duration of this test. Therefore, BSPs shall not require sensors or associated device drivers to be connected or installed on the test PC. Also, BSPs should not attempt to initialize an unconnected sensor at any point during this test.

User Interface

As this is an automated off-line test, BSPs should never display a graphical user interface (GUI), messages, dialogs, or wait for user input at any point during this test.

Primary Finger Verification

Since this is an automated off-line test, the test application and the vendor BSPs will not prompt for either one of the two fingers in the enrolled SID-0002 BIR. To test both fingers from an SID-0002 enrollment, the test application will produce two separate BIRs at run-time; each containing one of the two fingers in both the primary and secondary templates of the resulting SID-0002 BIR. Therefore, vendor BSPs should only verify against the **primary finger** in the SID-0002 BIR during a **BioAPI_VerifyMatch** call.

Vendor Software Requirements – Option 2 (Stand alone software)

To participate in off-line testing, vendors will need to provide two command-line applications to perform off-line SID-0002 biometric enrollment and verification on the Windows XP platform.

sidenroll.exe

This application will accept two raw BMP fingerprint images native to that product, process them for enrollment, and output an SID-0002 compliant enrollment BIR.

sidenroll /i1 *pathname* /f1 *finger* /i2 *pathname* /f2 *finger* /b *pathname*

Required Parameters

/i1 Pathname to an existing BMP image (Primary Finger).
This parameter shall contain a pathname to a raw BMP image file output by the same product during the 2004 ILO SID Biometric Technology Test.

/f1 Number from 1 to 10 to be used in the Finger Position field of the primary template of the SID-0002 BIR.

/i2 Pathname to an existing BMP image (Secondary Finger).
This parameter shall contain a pathname to a raw BMP image file output by the same product during the 2004 ILO SID Biometric Technology Test.

/f2 Number from 1 to 10 to be used in the Finger Position field of the secondary template of the SID-0002 BIR.

/b Pathname to a new binary file to contain the SID-0002 BIR.
This parameter shall contain the pathname of the file in which the new SID-0002 compliant enrollment BIR will be stored. The primary and secondary finger templates in the BIR will correspond to the fingers from /i1 and /i2 respectively.

Example

```
sidenroll /i1 c:\ProdA\112_2.bmp /f1 2 /i2 c:\ProdA\112_7.bmp /f2 7 /b
c:\ProdA\112.bir
```

Remarks

If either of the two fingerprint images is of insufficient quality for enrollment, the BSP should not create a BIR file. When the test application attempts to retrieve this (missing) file, it will record a failure-to-enroll for that seafarer and product.

sidverify.exe

This application will accept a raw BMP fingerprint image native to that product, attempt to match it against the primary or secondary finger in an SID-0002 BIR, and output the score and match result to an ASCII text file.

sidverify /i *pathname* **/b** *pathname* **/t pri | sec /r** *pathname*

Required Parameters

/i	Pathname to an existing BMP image. This parameter shall contain a pathname to a raw BMP image file output by the same product during the 2004 ILO SID Biometric Technology Test.
/b	Pathname to an existing SID-0002 BIR. This parameter shall contain a pathname to an SID-0002 compliant BIR containing two fingerprint templates.
/t pri | sec	A value of **pri** (primary) or **sec** (secondary) specifies which one of the two fingerprint templates in the BIR should be matched with the image provided.
/r	Pathname to a new text file to contain the match result and achieved FAR.

Example

```
sidverify /i c:\ProdA\112_2v4.bmp /b c:\ProdB\112.bir /t pri /r c:\test\result.txt
```

Remarks on Verification

The ASCII text file that is created to contain the match result and achieved FAR should only contain these two values separated by a comma.

Match
The first value should be 1 (Match) or 0 (No Match) which specifies whether verification was successful or not according to the internal threshold set by the vendor.

FARAchieved
The second value should be a FAR between 0 and 2^{31} -1 (2,147,483,647) indicating how closely the fingers actually matched, with 0 indicating a perfect match and 2^{31} -1 indicating a perfect non-match.

Sample contents of result.txt:
0,1332000123

General Remarks

System Stability
All vendor executables must coexist on a single client PC, so each vendor must mitigate against memory leaks or other issues that could cause instability of the test system after their executables have been called.

Unconnected Hardware
No sensors will be connected or used for the duration of this test. Therefore, executables shall not require sensors or associated device drivers to be connected or installed on the test PC. Also, the executables should not attempt to initialize an unconnected sensor at any point during this test.

User Interface
As this is an automated off-line test, the executables should never display a graphical user interface (GUI), messages, dialogs, or wait for user input at any point during this test.

Test Methodology

Phase 1: Off-line Enrollment

Raw image pairs of primary and secondary fingers of individual seafarers that were output by each vendor's BSP during onboard enrollment will be used to create new two-finger SID-0002 BIRs. These images were all captured with the sensor corresponding to the BSP that output the images.

Phase 2: Off-line Verification

Raw images output from each product during onboard verification visits 1, 2, and 3 for each seafarer will be matched against SID-0002 compliant enrollment templates created in Phase 1 for:

- the same seafarer on the same product (Native False Rejects)
- at least one other seafarer on the same product (Native False Accepts)
- the same seafarer on all other products (Non-Native False Rejects)
- at least one other seafarer on all other products (Non-Native False Accepts)

Simulated Attempts

All single-finger verifications will be executed three times by the test application; each time using a different raw image (when available) from the same seafarer and visit to simulate placement attempts. Only the best matching score (lowest value of FARAchieved) for all three simulated placements will be considered in determining whether a match or non-match was achieved for that finger at that threshold.