INTERNATIONAL LABOUR ORGANIZATION

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

ILO Seafarers' Identity Documents Biometric Interoperability Test (ISBIT-4) Report

Geneva, 2009

INTERNATIONAL LABOUR OFFICE GENEVA

978-92-2-123172-1 (print)
978-92-2-123173-8 (web pdf)

ILO publications and electronic products can be obtained through major booksellers or ILO local offices in many countries, or direct from ILO Publications, International Labour Office, CH-1211 Geneva 22, Switzerland. Catalogues or lists of new publications are available free of charge from the above address, or by email: pubvente@ilo.org

Visit our website: www.ilo.org/publns

Printed in Switzerland

INTERNATIONAL LABOUR OFFICE (ILO)
BUREAU INTERNATIONAL DU TRAVAIL
OFICINA INTERNATIONAL DEL TRABAJO

4, route des Morillons, CH-1211 GENEVE 22
Telephone +4122 799 61 11 Fac-simile +4122 799 86 85
E-mail: ilo@ilo.org
Telegramme INTERLAB GENEVE

ILO **S**eafarers' Identity Documents **B**iometric **I**nteroperability **T**est (**ISBIT-4**) Report

Second Revision

Principal Authors: Dr. John W. M. Campbell
john@bionbiometrics.com

Matthew J. Madden
matt@bionbiometrics.com

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). This revision of an 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, which is already underway in several countries, requires an internationally interoperable biometric to be used for verification of seafarer identities. In March 2004, the ILO Governing Body adopted the technical standard, ILO SID-0002 Finger Minutiae-Based Biometric Profile for the Seafarers' Identity Documents, as "The standard for the biometric template required by the Convention". This document defined the standard for the use of fingerprint minutiae templates as the interoperable biometric for SIDs. It 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 a biometric testing campaign (ISBIT-1) to develop a list of conformant and interoperable products for Members to use when implementing ILO Convention No. 185.

ISBIT-1 took place on board a cruise ship, the Crystal Harmony, and involved crew members from a variety of occupational and demographic groups. Of the seven products tested on board the ship, only two were able to interoperate at the ILO mandated performance level of 1% FRR at a 1% FAR. The experimental procedures, results, and analysis are described in the document listed on the ILO website at http://www.ilo.org/public/english/dialogue/sector/sectors/mariti/security.htm as, Biometric Testing Campaign Report (Part 1).

The interoperability of the tested products varied considerably and a study was launched to try and understand the causes of the lack of interoperability. One cause appeared to be some difficulty in interpreting the requirements of ILO SID-0002 and of the underlying ISO standards. Therefore an amended version of ILO SID-0002 was developed to provide additional emphasis on key areas, as well as to correct some minor errors in the original document. This amendment was approved by the ILO Governing Body in November 2005, and the version referenced above and available on the ILO website reflects these changes. After vendors revised their minutiae encoding and matching algorithms to reflect the new insights into key interoperability issues, six of the seven products were retested in an offline test using new algorithms with the fingerprint images from the original ISBIT-1. In this test, ISBIT-2, interoperability was substantially improved in all cases, but there were still only three products that achieved the ILO mandated performance level. The procedures, results, and analysis for this second test are described in the document listed on the ILO website as Biometric Testing Campaign Report (Addendum to Part 1).

In 2006, with the changes to ILO SID-0002 being formally approved and with the ISO standards that were referenced in ILO-SID being formally published, there was renewed interest in testing multiple products to try and achieve a larger number of products that could interoperate at the ILO mandated performance level and thus give ILO Members greater choice when selecting biometric products to use. A third test took place between January and June of 2006, this one in a biometric testing lab in Ottawa, Canada with a group of test subjects recruited from the local area. Although the test subjects and test environment were different in this case than in the previous test performed on board the ship, the presence of the three previously approved products allowed normalization between the results from the two tests. The improvements resulting from the maturity of the ISO standards and of the revised version of ILO SID-0002 were significant and a total of nine products were now demonstrated to be interoperable with one another.

This document describes a fourth test, which also took place at the testing laboratory in Ottawa, using a large portion of the same test crew and an environment as similar as possible to the third test. This test took place because of interest on the part of additional vendors to have their products tested for interoperability. It uses a methodology and report structure that is substantially the same as that of the third test, but includes results from a total of twelve biometric products.

Executive Summary

The ILO Seafarers' Identity Documents Biometric Interoperability Test #4 (ISBIT-4) took place in Ottawa, Canada from August to November, 2008. The nine products from the existing ILO list of approved products were tested together with three new products. Each product consisted of a sensor paired with an algorithm capable of both biometric enrolment and verification. There was also interest from two vendors in submitting updates to their existing fingerprint minutiae template generation and matching algorithms. Since these changes did not require any changes to the sensor or acquisition software, it was decided to treat these independently from the primary test. The results of these algorithm changes were generated by reprocessing the test data during December, 2008 and January, 2009 and are shown in Annex D of this report.

The initial phase of conformance and basic interoperability revealed that all twelve products were conformant to the requirements of ILO SID-0002 and could achieve interoperability with one another on a limited set of fingers. Preliminary testing and integration of all the products with the test control software took place mainly in August and September of 2008.

Full data collection from 189 test subjects aged from 18 to 69 took place in September and October of 2008. Each test subject visited the test laboratory twice during that period and was enrolled on each product and verified multiple times on each product during each visit. A total of 84,806 fingerprint images were collected under controlled and supervised conditions.

Data processing and analysis took place in November, 2008 with the final production of this report in June, 2008. A total of 119,197,971 individual matches were computed, resulting in a total of 20,251,227 two finger transactions being simulated.

Of the new products tested, all three achieved the target interoperable performance metric when used in conjunction with the previously approved products and with each other of a mean GFRR of 1% or less at a GFAR of 1% (See Section 4.1 for definitions). Specifically the mean GFRR was 0.80%. One product did produce two non-conformant templates during the test but the vendor agreed to modify the algorithm to correct this error and include it to be retested with the other two previously approved algorithms that were to be modified. When the modified algorithms were substituted and the offline matching was reprocessed, these twelve products also met the ILO performance target, with a mean GFRR of 0.80%. None of these products produced any non-conformant templates. One of the modified versions of a previously approved algorithm, however, had a maximum GFRR of 3.7% which was greater than the maximum GFRR when using the previous algorithm version from the same product. It therefore can't be recommended as a replacement for the previously approved algorithm.

The overall recommendation is that the approved products list be amended to include the three new products and one modified product as noted in the proposed list of approved products provided in a separate attachment with this report.

Contents

Index of Tables

Index of Figures

1 Introduction

1.1 Background

Fingerprint matching based on minutiae points is a well established technology in the biometrics industry. Historically each vendor used an internal minutiae format 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 generally be correctly interpreted by other vendors' products and so fingerprint enrolment and verification had to use equipment supplied by the same vendor. More recently the International Organization for Standardization (ISO) has published a number of fingerprint data interchange format standards intended to enhance interoperability. One of these, ISO/IEC 19794-2:2005 Biometric Data Interchange Formats - Part 2: Finger Minutiae Data, defines the content and structure of standardized fingerprint minutiae templates at a variety of compression levels, including some designed for use in identity cards with limited data storage space.

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 seafarers' identities and their associated entitlement to the special privileges related to transit through or entry into countries (such as for shore leave) that the Convention grants to 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 March, 2004, the ILO approved ILO SID-0002 Finger Minutiae-Based Biometric Profile for the Seafarers' Identity Documents as "The standard for the biometric template required by the Convention". This document provides details about the biometric template and other data to be stored in the bar code and also defines appropriate methods for enrolment and verification of the seafarers' fingerprints. The fingerprint minutiae standard used as the basis of the biometric template was a preliminary draft of the SC 37 standard ISO/IEC 19794-2:2005 referenced above and the specific format selected was the normal sized finger minutiae card format with headers. ILO SID-0002 was subsequently updated based upon the results of the ISBIT-1 and ISBIT-2 tests described below, and the revised document (which is the one provided in the link above) was approved in November, 2005.

1.2 ISBIT-1 and ISBIT-2

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. Multiple biometric products (each consisting of a fingerprint sensor combined with an enrolment and verification algorithm) were submitted by different vendors, to determine whether or not the products could achieve conformance to the standard and, if conformant, could achieve 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.

The initial lab testing for conformance found seven products that were declared conformant and thus suitable for the full test, which took place in September and October of 2004. In order to simulate operational enrolment and verification of seafarers in a realistic environment, all the tests involved live capture of fingerprints from seafarers. The detailed test methodology employed and the results obtained are described in the ILO Biometric Testing Campaign Report (Part 1) which can be found at the URL http://www.ilo.org/public/english/dialogue/sector/sectors/mariti/security.htm (the ILO maritime security website). Due to the requirement to support live capture on multiple products, 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 26,088 live single-finger verification transactions, each consisting of up to 3 single finger presentations. This test is known as **I**LO **S**eafarers' **I**dentity Documents **B**iometric **I**nteroperability **T**est #1 or ISBIT-1.

The results of ISBIT-1 were mixed. Some products performed very well with fingerprint templates from all of the participating vendors. Others performed poorly with all templates except the ones they had produced. Still others performed well with templates from some products and poorly with templates from other products. Clearly there was an interoperability problem. Given that the ISO fingerprint standard being implemented was a draft and none of the companies had previous experience implementing this standard, these results were not surprising. After extensive discussions with the companies on the possible sources of interoperability problems and a careful review of the draft standard, a supplementary guidance document was produced to aid the companies in achieving interoperability.

Six of the seven original products had their algorithms updated to reflect the guidance provided in the interoperability document (one vendor elected not to participate) and were then used in a second test. In this test, ISBIT-2, the live capture transactions were simulated using the images that had been stored during the previous test. The results showed that interoperability was substantially improved in all cases. Due to the stringency of the ILO performance requirement of 1% or lower FRR at a 1% FAR, however, there were only three products that achieved the ILO mandated performance level and were published in a list of approved interoperable products. The procedures, results, and analysis for ISBIT-2 are described in the document listed on the ILO website as Biometric Testing Campaign Report (Addendum to Part 1).

1.3 ISBIT-3

In 2005, after the completion of the original ISBIT-1 and ISBIT-2 tests, ISO published the final version of ISO/IEC 19794-2. There were also a number of other programmes beyond ILO SID which began to require standards based fingerprint minutiae matching. Due to this increased interest in fingerprint minutiae interoperability, there were more vendors spending more effort on developing conformant and interoperable products. By the end of 2005 many additional vendors wished to submit products to be tested for inclusion on the ILO list of approved products. The ILO Seafarers' Identity Documents Biometric Interoperability Test #3

(ISBIT-3) was therefore commissioned in late 2005 and took place from January to June of 2006.

The goal of ISBIT-3 was to combine the best characteristics of ISBIT-1 (live capture under controlled conditions, direct feedback to the participants of whether each placement matched or not) and ISBIT-2 (ability to generate artificial transactions using combinations of pre-acquired images and thus generate reliable imposter match statistics for each combination of enrol and verify product). Due to cost constraints, ISBIT-3 used a test lab in Ottawa, Canada with a test crew recruited locally rather than a group of seafarers on board a ship. Each test subject visited the test lab twice, with enrolment and verification on each product during each visit. The enrolment session from the first visit and the verification session from the second visit (known as E1V2) were used together, as were the enrolment session for the second visit and the verification session from the first visit (known as E2V1). The idea was to get twice as much data with enrolment and verification separated by two to three weeks, even if in one case enrolment was before verification and in another, enrolment was after verification. The user feedback of match or non-match for a given finger placement was always based on the enrolment for the first session which meant a slight difference of user feedback between the two visits but the final results showed that this made no statistically significant difference in the performance measured using the E1V2 data versus that measured using the E2V1 data.

This test also introduced a method of performance normalization that was designed to account for any differences introduced by the change in environment or the change in test crew between this test and the previous test. Specifically, the test included the three existing approved products as well as six new products from six new biometric vendors and the performance criterion that had to be met in order for a new product to be added to the interoperable products list was modified to normalize differences between the tests. The error rates were also computed using a more sophisticated generalized transactional false accept rate (GFAR) and generalized transactional false reject rate (GFRR), which accounted for the effects of failures to enrol, failures to acquire, and the way in which the ILO operational enrolment and verification transactions were specified in ILO SID-0002.

The refined metrics for a group of products to be approved as meeting the ILO performance criterion were therefore define such that the mean GFRR at a GFAR of 1% for all of the products on the new approved list had to be less than or equal to the maximum of:

- 1%, and
- the mean GFRR at a GFAR of 1% for the three original products using the data collected during ISBIT-3

This allowed for the possibility that the new test might involve a more difficult demographic group or more difficult environmental conditions than the original test, which would force the mean FRR to be higher, even for the approved products. It also meant that if the test turned out to be particularly easy for the approved products, the threshold for qualification still would not be set lower than the original 1% mean criterion established for ISBIT-1 and ISBIT-2.

The results of ISBIT-3 were very encouraging. Interoperable performance had obviously improved substantially since ISBIT-2 and all six of the new products were added to the approved products list. In fact, it turned out that the new products performed better on average than the old products and if it not been mandatory to retain all three of the previously approved products then performance could have been improved still further.

1.4 Current Test

After the completion of ISBIT-3 and the publication of the new list of approved products by ILO in November, 2006, there was a temporary pause in testing activity. Those biometric vendors that had participated in the test were satisfied with their results and there were a reasonable variety of products for ILO member states to choose from. By late 2007, however, there was interest from some new vendors and once there were three products ready to be tested, the interest was deemed sufficient to mount another test. The cost of the test grows as new products are added since each new test requires data to be acquired not only from the new products but also from all of the previously approved products so that the test of interoperability involves the same test crew under the same environmental conditions. Three products was an absolute minimum to justify a new test and therefore the new test did not begin until all three were ready in August, 2008. This ILO Seafarers' Identity Document Biometric Interoperability Test #4 (ISBIT-4) took place from August to December of 2008 and is the subject of the present report.

A total of twelve products were tested, using the same methodology as in the previous ISBIT-3 test, although the software that controls the data acquisition and matching (the test harness) was refined to make testing more efficient and an additional quality control step was added to try and eliminate any human errors in the database. This was not done in previous tests because it was very labour intensive and there was already close supervision of every fingerprint capture as well as detailed computer and human prompts to instruct the test crew. Despite the test crew also being prompted and supervised in ISBIT-4, however, the quality control step was found to be beneficial as it identified various errors that had somehow slipped through. It is likely that similar errors existed in the previous tests, but the number of errors observed was so small (< 0.1%) that it would not have affected the outcome of the previous tests.

A test crew of 200 people aged from 18 to 69 was recruited. Some did not complete both visits and were removed from the test, but a total of 189 were present for both visits. Of these, 125 were from the previous test crew used in ISBIT-3. Since that test crew only included 184 subjects between 18 and 69 and since a few of those passed their 70th birthday between tests, this is a good repeat rate between tests that are separated by over two years in time. It means that the test crew for ISBIT-4 overlapped that in ISBIT-3 by 68%, reducing any performance differences due to changes in test crew.

One change was made to the interoperable performance criterion established in the previous tests. This occurred because of the comments from independent experts who had reviewed the ISBIT-3 test and noted that as the number of approved products grew, it was possible for a new product to perform poorly but for the superior performance of the previously approved products to keep the overall performance of the group including the new product below the ILO performance threshold defined by a mean GFRR of 1%. This was the drawback to using the mean GFRR as a criterion since outliers could affect the operational use of the global SID system without significantly affecting the mean in the test. The performance criterion was therefore modified to have two parts which both had to be satisfied for new products to be added. This meant that the fundamental performance criterion did not change, but the second criterion prevented a poorly interoperable product from being added to the list simply because it did not sufficiently perturb the mean.

1) The mean GFRR at a GFAR of 1% for all of the products on the new approved list had to be less than or equal to the maximum of:
    - 1%, and
    - the mean GFRR at a GFAR of 1% for the nine previously approved products using the data collected during ISBIT-4
2) The maximum GFRR at a GFAR of 1% for all combinations of enrol and verify products in the new approved list could not be larger than the maximum GFRR at a GFAR of 1% for all combinations of enrol and verify products among the nine previously approved products

Due to the increased number of products, ISBIT-4 included even more fingerprint images than ISBIT-3. The total number of matches and transactional matches computed is less, however, because ISBIT-4 did not process matches for enrolment and verification combinations that occurred during the same visit. These were processed in ISBIT-3 although they were not used in computing the final interoperability matrix that determined the list of approved products. The relative sizes of all the tests thus far are shown in Table 1 below.

Table 1. Relative Size of ISBIT-1, 2, 3 and 4

Test Name	Number Of Products	Test Crew (Number of People)	Finger Images Collected	Number of Two finger Transactions	Number of Single-Finger Single Presentation Match Attempts
ISBIT-1	7	126	26,948	13,044	26,067
ISBIT-2	6	126	26,948	67,307	403,844
ISBIT-3	10[1]	191[2]	67,802	27,066,803	161,359,702
ISBIT-4	12	189	84,806	20,251,227	119,197,971

[1] In ISBIT-3, one product had two variants tested. This resulted in data capture for 9 products but matches computed for 10 products

[2] In ISBIT-3 the total test crew was 191 subjects, but 7 of them were 70 or older. This seemed unlikely for active seafarers and so the test results were based on the 184 subjects aged 18 to 69. In ISBIT-4, no test subjects older than 69 were recruited.

2 Test Methodology

2.1 General Test Conditions

2.1.1 Environment

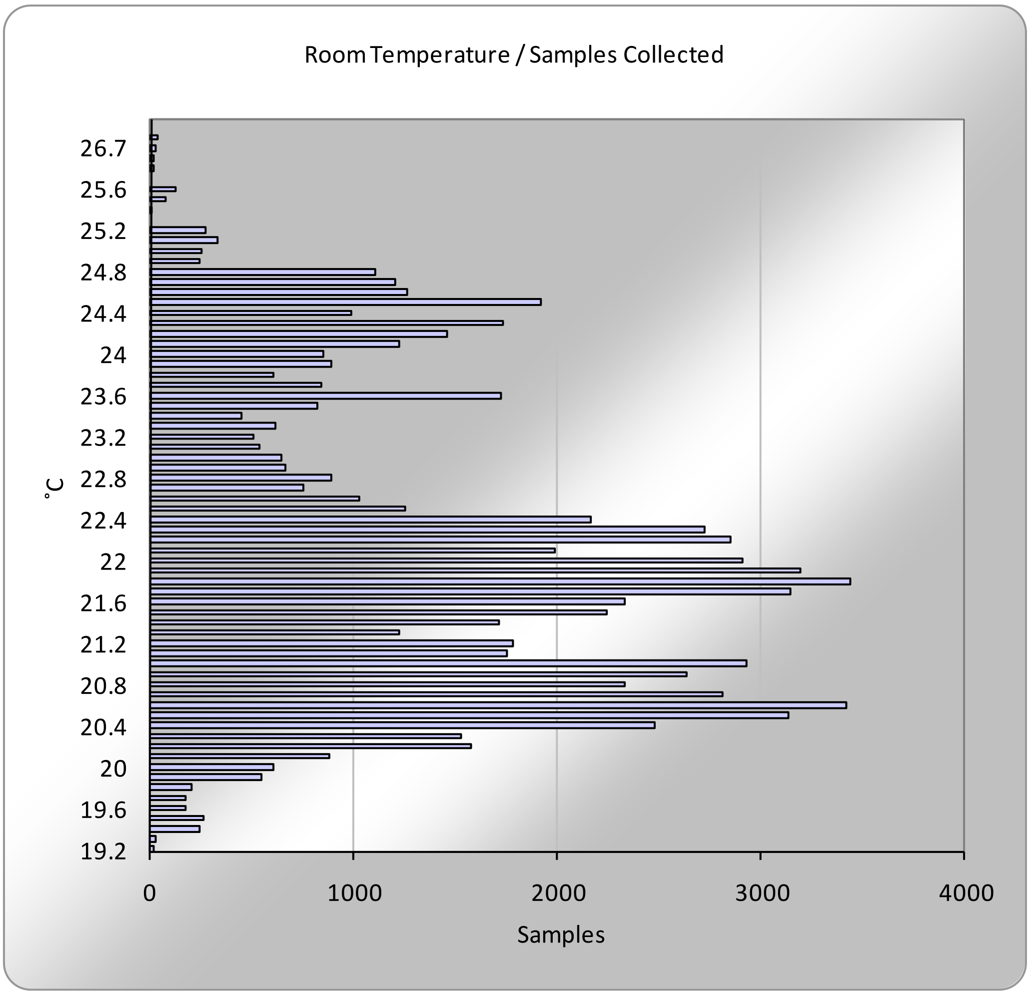
Live fingerprint capture from the test subjects occurred in a "normal office environment," under indirect fluorescent lighting, during the months of September and October, 2008 in Ottawa, Canada. Unlike ISBIT-3, which took place during the Winter and required the use of a humidifier to keep the relative humidity above 20%, there was no need for specific humidity control beyond that provided by the office building heating and air conditioning system.

During data acquisition, room temperature and relative humidity were sampled at two minute intervals using the Extech Instruments Temperature / Humidity Datalogger 42270. By correlating this data with the timestamp of all fingerprint samples collected we found the following:

Temperature

The mean temperature for all samples collected was 22.0 ˚C, while the minimum was 19.2 ˚C, and the maximum was 27.1 ˚C. The small number of samples at an anomalously high temperature was probably caused by the Extech Datalogger being accidentally placed near the exhaust port of one of the computers one morning, which caused a temporary spike in the recorded temperature.

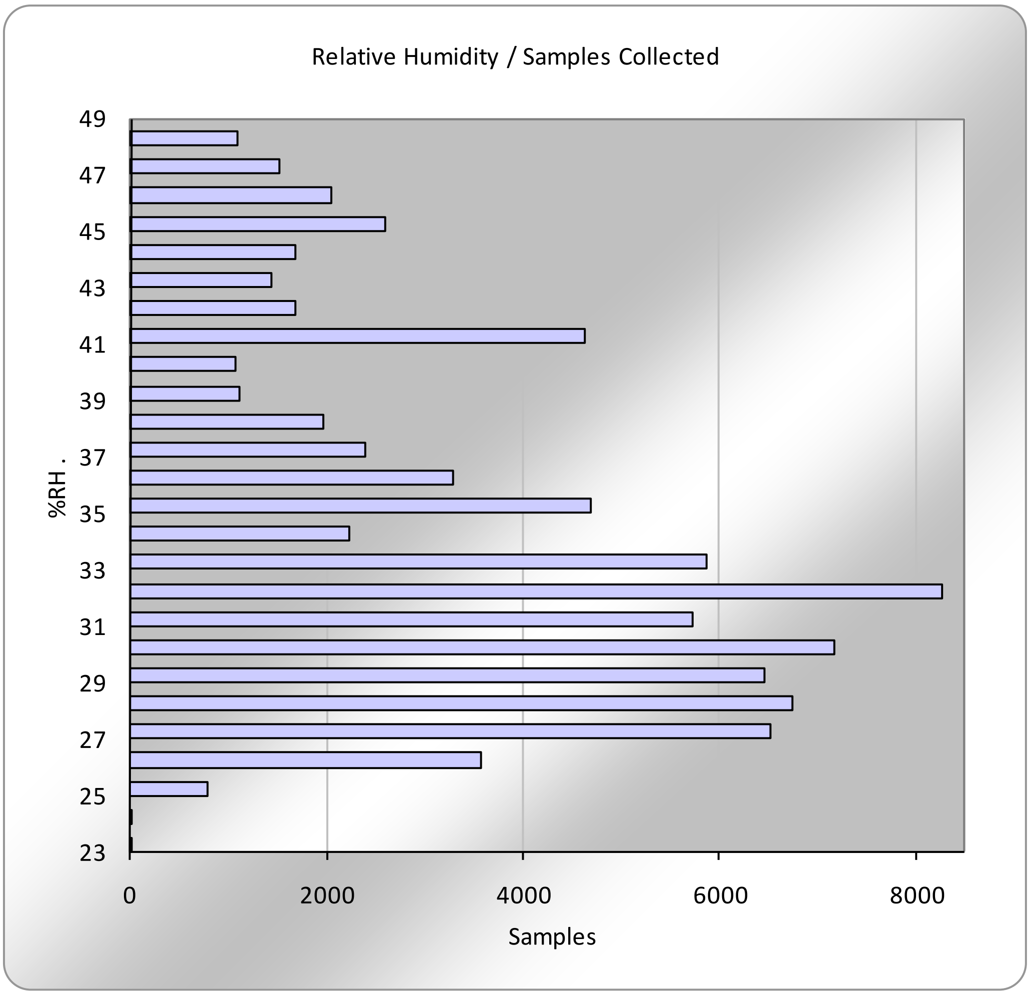
Figure 1. Samples Collected by Room Temperature



Humidity

The mean humidity for all samples collected was 34% RH, while the minimum was 23% RH, and the maximum was 49% RH. This compares with a mean humidity of 30% RH, a minimum of 13% RH, and a maximum of 42% RH for ISBIT-3, suggesting that excessively dry air and associated dry fingers was not as significant a problem in ISBIT-4 as it was in ISBIT-3.

Figure 2. Samples Collected by Relative Humidity



2.1.2 Sensor Maintenance

Data acquisition generally took place from 9:00 AM until 9:00 PM. During an appropriate break in the early afternoon and again at the end of the day, each biometric sensor was wiped clean with an alcohol swab. In between these formal cleanings individual sensors were wiped with a dry cloth or with an alcohol swab whenever a problem with dirt or residue on the sensor surface was observed. The specific instructions to the test crew regarding sensor maintenance are included in as part of the overall instructions to the test crew shown in Annex C of this report.

2.1.3 Order Effects

The order in which the biometric products are used can potentially affect performance for a variety of reasons such as the following:

- Feedback from one biometric product may affect test subject behaviour (e.g. finger pressure) on another.
- As each product is used, the test subject becomes increasingly habituated to presenting their fingerprint and thus may achieve better results with later products.

- If a test subject has difficult fingers, it may take them some time to learn the best strategies to cope with this (moisturizing, drying fingers, etc.) and therefore the first few products they use may capture lower quality fingerprints.

In an effort to eliminate all order effects, the test harness randomly selected the order of products for the enrolment and verification phase of each visit, with a different random order for each phase. The software did not allow any deviations from this pre-selected random order and there were in fact two occasions where sensors failed and the visit had to be temporarily delayed while they were repaired.

2.1.4 Product Solicitation and Integration

All vendors expressing interest in participating in the test were provided with two documents in advance of the deadline to deliver products to the test lab. The first, contained in Annex A of this document, was the test API specification required to integrate their product with the test harness. The second (contained in Annex B of this document) was the test methodology which explained to the vendor exactly how their product would be tested and what was the criterion for it to be accepted for addition to the ILO list of approved products. Interested vendors were required to provide two copies of their sensor and five copies of their algorithm to the test. Integration of the products into the test control software then took place at the same time as conformance testing to determine if each product was conformant to the requirements of ILO SID-0002 and of the test API. During August and early September, any issues related to conformance or integration were reported to the vendors and resolved before the final test phase, when capturing fingerprints from the test crew began. During this integration phase, there were no formal rounds of information exchange between the test lab and the vendors, since any problem was reported to the supplier of a particular product as soon as it was detected and the product was retested as soon as they provided a fix. This phase ended on September 26, 2008.

2.1.5 Test Team

During the data capture phase the test team consisted of three members: an experimenter, and two test administrators.

The experimenter was responsible for the overall design and management of the test, ensuring consistency in the guidance provided to the test subjects, and reviewing test results on an ongoing basis to ensure integrity. In particular, the experimenter was responsible for the quality control phase of data acquisition. This involved a manual review from the test database of every fingerprint image captured by every product for each test subject at each visit.

The administrators were responsible for operating the test harness and supervising the test subjects. This began with ensuring that the test subjects correctly filled out their data release forms and verifying that their demographic information was correctly entered in both the paper forms and in the test harness. It included providing each product for the test subject from the line of twelve products at a single test station when the test harness prompted for it and providing appropriate guidance to each test subject about the purpose of the test, the difference between enrolment and verification, and how to place their finger on a fingerprint sensor. It also included providing suggestions if the test subject had problems with fingers that were too moist, too dry, or too dirty. Perhaps the most important job of the administrators was to ensure that the test subjects actually followed the instructions of the test harness in terms of when to present a finger to a given product and especially of which finger to present.

Since the ILO SID is expected to be used operationally under supervised enrolment (at SID enrolment centres) and verification (at ports or at border crossing points), the supervision by an administrator was required to simulate operational conditions. It was also required to help the test subjects, who were often initially unfamiliar with fingerprint technology, become rapidly familiar with the correct way to present a finger to a given product.

The administrator also has the role of correcting any minor malfunctions with the sensors or software and with escalating such malfunctions to the experimenter or even to a member of the test harness development team if they could not be quickly resolved. The general instructions provided to the administrators are shown in Annex C of this report.

2.1.6 Test Harness and Testing Stations

The test laboratory had two testing stations, each consisting of three computers with a single monitor, keyboard and mouse. The twelve fingerprint sensors and associated software that comprised the products being tested were distributed among the three computers. The test harness had a client component that was installed on each computer and managed the interface with the product software provided by each vendor and the feedback to the test administrator and to the test subject. The test harness also had a server based database component that recorded all the data and managed such issues as the order of the components, which products were installed on which client computers and similar issues.

The reason that three computers had to be used at each test station was that many of the products had conflicts with each other and this was the minimum number of computers that could support the products without catastrophic failures. Even with three computers there were still multiple products and the test harness itself installed on each computer. Although the integration testing phase had ensured that each product could work individually with the test harness and properly supported the test API and had allowed the major incompatibilities among products to be determined, there were still some ongoing issues that were very difficult to address. Over the last two weeks of the integration phase and the first two weeks of the data acquisition phase of the test, these were identified to be problems with memory and resource management in Windows. Even among those previously approved products that had been able to work reliably together in ISBIT-3, new compatibility problems arose. These would cause periodic crashes of either the product software itself or sometimes of IsbitCapture and IsbitDirector, the two main components of the test harness. Since the core modules of the test harness that interfaced with the products being tested had not changed since ISBIT-3 and neither had the previously approved products, the most likely reason for the new compatibility problems was the updates in the Windows operating system and the associated software such as the Microsoft .NET Framework.

One consequence of this for the test itself was that in the first few days, when the number of crashes was highest, there was one test subject who was unable to complete his first visit before having to leave to attend an external appointment. Although this person subsequently completed the visit, it was many hours later and so the process they followed was different than that of all other test subjects. This test subject was therefore deleted from the database and is not included in the 189 test subjects considered in this report.

As testing progressed, solutions were developed that encapsulated the code from the different products in different ways. By the third week of data acquisition these errors had been almost completely eliminated. Prior to that, procedures had been put in place for the test administrators to deal with these errors and the test harness was sufficiently robust that in almost all cases these errors did not affect the data acquisition other than to introduce a small delay between fingerprint presentations as the error was dealt with at the appropriate level. The detailed instructions to the test administrators on how to deal with these errors are included in Annex C of this report. There were a few cases at the beginning of the test when the errors were not properly handled by the test harness, resulting in incorrect transaction metadata being associated with certain fingerprint images in the database. This resulted in a particular product at a particular visit having transactions that consisted of either more or less than three fingerprint presentations for a single finger. Since the test harness only accepted transactions containing three fingerprint presentations of each finger, it was very easy to identify these errors and manually correct them by adjusting the metadata. In total, only 12 transactions of all those acquired during the data acquisition had to be modified.

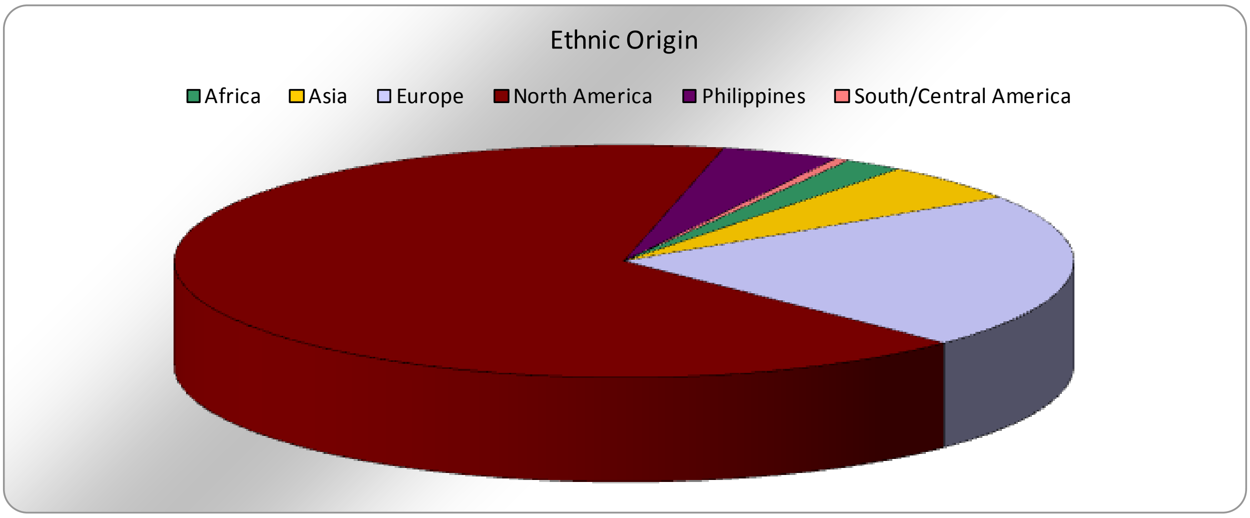
Another issue that occasionally occurred was that a sensor would be accidentally disconnected from a computer as it was moved from the line of sensors at the testing station and placed in front of the test subject when its turn came in the random enrolment or verification order. In this case, the sensor was reconnected and then the driver would sometimes have to be reinstalled. Sometimes the drivers also had to be reinstalled when the computers were rebooted. Each test administrator was given instructions on driver reinstallation and this did not result in any significant delays.

The final error that occurred was that on two occasions a sensor failed and needed to be fixed by having its flash memory rewritten. This did cause a delay in the visit for the test subject involved, but had no other adverse effect.

2.1.7 Test Crew

The test crew was made up of 189 volunteers from the Ottawa area that were willing to submit their fingerprints and provide some limited demographic information as part of being tested. They were aware of the purpose of the test prior to their participation, and were required to sign a personal data release form. The ethnic origin of this test crew was biased to North American individuals, but contained a mixture of individuals from other parts of the world, as shown in the figure below.

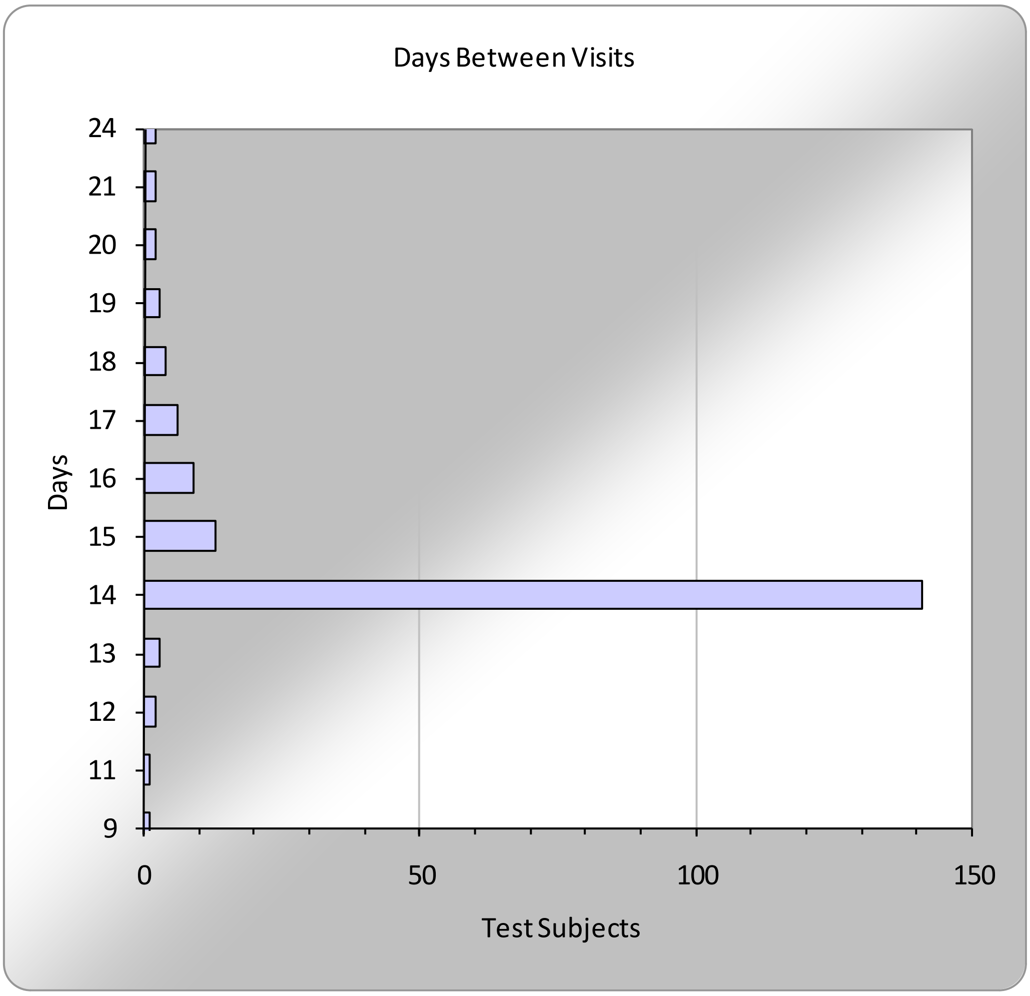
Figure 3. Ethnic Origin of Test Crew



The test subjects were asked about the degree (None, Light, or Heavy) to which their work or in some cases their hobbies involved manual or chemical exposure to their fingers. Since each test subject had varying opinions about this, some guidance was given based on past experience from the previous ILO tests. Nurses, for instance, tended to have heavy chemical exposure, especially if they worked in a facility that used an alcohol based cleanser instead of soap for washing hands. Construction labourers also were marked as heavy. Elementary school teachers were marked as light because of frequent hand washing. Office workers were marked as none. Overall, 16% of the test crew indicated heavy manual or chemical exposure, 30% indicated light exposure, while the remaining 54% indicated no exposure.

Each test subject made two visits to the test lab for the online component of the performance and interoperability test phase. Reasonable effort was made to schedule the visits approximately two weeks apart. The actual mean duration between visits was 14.6 days, with a minimum of 9 days and a maximum of 24 days. The distribution of the time between visits is shown in Figure 4 below.

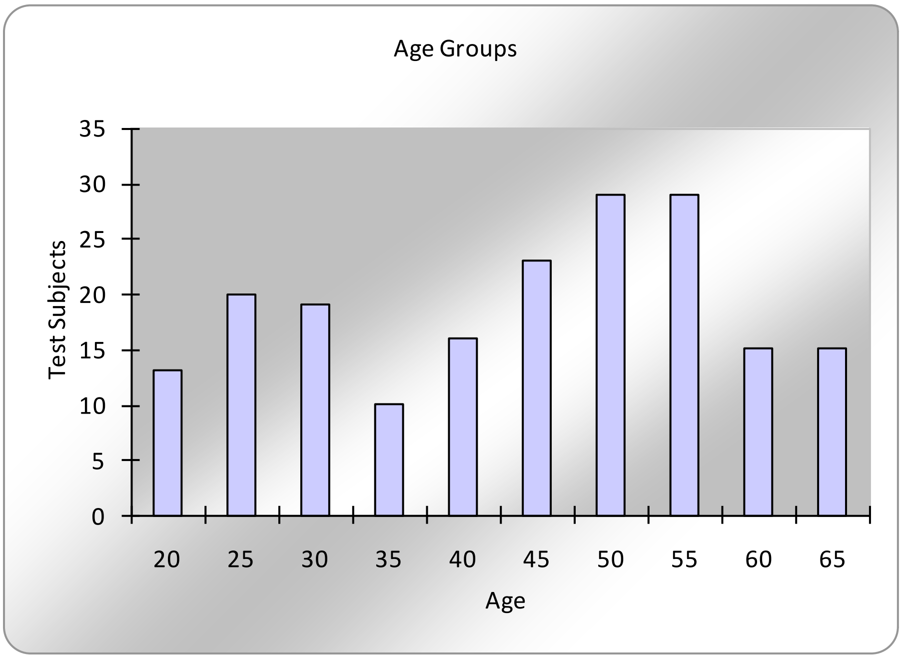
Figure 4. Distribution of Days Between Visit 1 and Visit 2



.

The test subjects were also widely distributed in age, with a range from 18 to 69 years old at the time of the test. The histogram below shows the age distribution of the test subjects.

Figure 5. Age Distribution of Test Crew



Each test subject received both verbal instructions from the test administrator and visual instructions from the test harness and any data capture GUI associated with each product. They were instructed when to place a finger, and when to remove it. If they placed their finger on the fingerprint sensor during calibration or if they removed it before the capture process was complete, then the administrator would inform them and activate a radio button on the test harness GUI to force the calibration or fingerprint presentation to be repeated. The capture process for a single finger presentation was considered to be complete when either a) the biometric product indicated a successful capture, or b) the biometric product indicated that it failed to acquire an image of acceptable quality or c) the 12 second timeout was reached before the biometric product returned any result. The test subjects also received guidance if their fingerprints obviously appeared to be of poor quality or if they were having difficulty in achieving consistent successful fingerprint captures. If the image appeared too light, the test subject was instructed to brush the finger along the side of the nose or forehead to moisten it and then the next placement was attempted. For individuals with poor ridge definition, or chronic dryness, moisturizing lotion was applied. The decision to apply moisturizer was always made within the first few capture attempts and the approved practice was to restart the visit after applying the moisturizer so that all products benefited equally from it. If the test subject had fingerprints that were too moist, they were asked to wipe the finger on a dry cloth or their clothing.

2.2 Performance and Interoperability

The objective of the Performance and Interoperability phase of ISBIT-4 was to determine both native (enrol and verify using the same product) and non-native (enrol using one product and verify using another) false reject and false accept rates for biometric verification of the test population over a reasonable period.

The performance component sought to demonstrate that the biometric products submitted for testing were able to provide sufficient accuracy to meet the ILO's requirements.

The interoperability component sought to determine the largest combined set of products which could achieve the ILO's requirements when working together, with enrolment on one product and verification on another.

2.2.1 Enrolment

Test subjects were enrolled on each biometric product during both visits in accordance with the requirements stated in ILO SID-0002. Any unavailable fingers (amputated, bandaged, etc.) were marked as such in the system and were not used in either the enrolment or verification phase for that user during that visit. During enrolment, a test subject made two enrolment attempts, each consisting of three single finger presentations, to enrol a primary and a secondary finger, starting with the right and left index fingers respectively. If an index finger was marked unavailable or produced such poor quality images that a successful enrolment was not possible, then the test team tried to enrol a fingerprint from another finger or thumb according to the order defined in ILO SID-0002, Section 5.1.1.

When none of the subject's ten fingers could be enrolled, then that test subject was recorded as being unable to enrol on that biometric product for that enrolment visit. The test subject was not able to participate in native genuine comparisons on that product during the online verification, although the test subject still participated in impostor comparisons and non-native genuine comparisons on that product.

All of the output images and biometric data interchange records (BDIRs) were stored in a secure database for subsequent online and offline verifications.

2.2.2 Online Verification

Immediately after each visit's enrolment session, the test subject made a limited number of genuine comparisons against a previously enrolled template on each biometric product. To maintain active participation by test subjects, the match/non-match decision for each presentation was displayed on the screen and read aloud by the administrator. In this way, online verification also functioned as a controlled data collection of images for all offline genuine and impostor comparisons.

The test harness software determined the unique finger positions enrolled by that test subject during that visit for all biometric products (usually the right and left index but sometimes additional fingers on certain products). Two single-finger attempts (each made up of three presentations) were captured on each biometric product for each unique finger position. Thus, if the test subject successfully enrolled their right and left index fingers on at least one biometric product, they presented each finger six times for a total of 12 verification presentations per product per visit. Alternatively, if they enrolled a left index and right thumb on one product, and right and left index fingers on all other products, they would have verified their right index, left index, and right thumb on all products for a total of 18 verification presentations per product per visit. This process was intended to ensure that images would be available for exhaustive native and non-native genuine comparisons offline.

The images that were captured during online verification were matched against BDIRs generated during the enrolment phase of the first visit. In addition to randomizing the order of the verification products, the test harness randomly selected a BDIR that was 80% of the time from the same test subject (genuine) on the same product (native) and 20% of the time from a different test subject (imposter) on the same product (native). In either case, if no BDIR containing the current finger was available from the same product then a genuine or imposter BDIR would be selected from another product (non-native).

Note that the manufacturers of the biometric products established initial threshold settings that were used for online verification, and these determined the match/non-match decisions provided as feedback to the users. The internal match decisions were not used in producing the results shown in this report. Instead, the results were based on the match scores produced during offline verification (described below) which were subsequently used to determine the optimal threshold settings for maximizing interoperability.

2.2.3 Offline Verification

Offline testing allowed exhaustive native and non-native genuine comparisons to be performed. That is, every presentation of a test subject's finger was matched against every BDIR containing the same finger enrolled by the same test subject on all biometric products. Normally this involved three presentations of each finger and the maximum similarity score of all three would be used as the similarity score for that attempt.

Similarly, exhaustive native and non-native impostor comparisons were performed offline by matching every verification image with every BDIR containing the same finger for all other enrolled test subjects on all biometric products.

Two finger verification transactions, as defined in ILO SID-0002, were simulated during offline testing by taking the maximum similarity score of each correlated pair of match attempts using the two fingers from each individual enrolment BDIR. If a BDIR contained only a single enrolled finger, then only a single match attempt was used to compute the transactional similarity score. If no match attempts existed for the corresponding primary or secondary fingers in the enrolment BDIR, a transaction was not simulated for that combination. If the fingers corresponding to those in the enrolment BDIR were marked as being unavailable during the online verification visit (because of physical damage or missing finger), and would have otherwise been genuine attempts, the transaction was simulated as a failure to acquire.

3 Conformance

Prior to being tested for performance based interoperability, products first had to be integrated into the test harness, a set of control software that facilitated test operations and data recording. In order to do this, they needed to comply with the requirements of the ISBIT-4 API Specification, as described in Annex A. This was provided to the vendors prior to the start of the test so that they could prepare appropriate versions of their software. Once these were ready, a series of conformance tests were performed to ensure that each product supported the API. If the product did not support the API, it could not be integrated into the test harness and it could not be tested. Since the API conformance requirements were really a requirement of the test and not of the eventual operational environment, the test lab did try and work closely with vendors to accommodate their products where possible. For example, under certain circumstances one of the previously approved products returned incorrectly formatted Windows Bitmap files, which could not be processed for enrolment or matching. This was not permitted by the API Specification, but was not a requirement of ILO Convention No. 185 or of ILO SID-0002. Therefore a workaround was added to the test harness to check for this and mark the files as impossible to process, so that they resulted in single presentation failures to acquire or failures to enrol.

The second, more critical, phase of conformance testing involved verifying that products supported the requirements of ILO SID-0002. This specified, among other things, the format of the minutiae based biometric data interchange record (BDIR) to be produced during a two finger enrolment and the specific means, when it was appropriate, of recording in this BDIR the fact that only a single finger or no finger at all could be enrolled. The formal mechanism for this is to use a special form of the minutiae template defined for an "unenrolled"[3] finger. The second portion of the conformance tests therefore involved a number of procedural tests to ensure that the right types of BDIRs were produced under different circumstances and that matches and non-matches occurred when they were supposed to (using some clear, high quality fingerprints) and that the data structure of the BDIRs produced in each test conformed to the requirements of ILO SID-0002. The tests in the second phase of conformance testing could be broken down into two categories.

3.1 Enrolment

Several enrolment trials were performed to ensure that each biometric product:

- prompted for placement of all ten possible finger positions by name
- provided visual feedback of the fingerprint image presented to the sensor
- indicated a failure to acquire or failure to enrol for fingerprints of insufficient quality
- successfully enrolled two fingers if two fingers of sufficient quality were available
- successfully enrolled one finger (in the event no other finger was available)
- produced BDIRs conformant to the data format specified in ILO SID-0002 Annex B

[3] "Unenrolled" fingers as defined by ILO SID-0002 Annex B (revised) are representations of fingers that are missing, damaged, or otherwise unable to be enrolled by a biometric system on the ILO approved products list.

3.2 Verification

Several verification trials were performed to ensure that each biometric product:

- prompted for placement of all ten possible finger positions by name
- provided visual feedback of the fingerprint image presented to the sensor
- indicated a failure to acquire for fingerprints of insufficient quality
- correctly interpreted conformant BDIRs containing both enrolled and "unenrolled" fingers
- indicated a match for genuine comparisons with some sample high quality fingers
- indicated a non-match for a selection of impostor comparisons
- provided a similarity score as defined in the API Specification

3.3 Basic Interoperability

If there were conformance problems in the first two phases then the vendor of the non-conformant product was allowed to try and resolve the problem. Once conformance was achieved by a product using only its own BDIRs, a basic interoperability test was performed to ensure that this product could successfully work with BDIRs produced by the other conformant products. This helped to reveal any subtle conformance issues. Sometimes, two products can both be conformant to a standard, but they may have chosen to write different values in a specific location of the data record where the standard allows flexibility. One of the products may be able to accommodate either value, but the other may not, resulting in a lack of interoperability.

A small number of fingerprints from a subset of the test crew were captured for enrolment and for verification on each product. The enrolment images were used to produce conformant BDIRs by each system that had passed the previous phase of conformance testing. The verification images were used to initiate genuine match transactions using the corresponding BDIRs produced by every product. In general, if the verification component of a product's software could read and produce match scores using one BDIR from a particular product then it would produce reasonable match scores for all of the BDIRs produced by that product (when matching against images acquired on its own sensor from the same test subject).

One lesson learned from ISBIT-3 was that multiple rounds of interaction with the vendors as they strived to achieve conformance and preliminary interoperability could be a lengthy process. In order to facilitate this phase, a web site was made available to participating vendors where they could download a set of sample enrolment and verification images along with the corresponding enrolment BDIRs for all of the nine previously approved products. They could also download a GUI based test application that allowed them to exercise all of the calls in their software API using the same calling code that would be used in the full test harness software. This gave them the ability to test some aspects of conformance and interoperability before sending their product to the test laboratory and it significantly shortened the effort that needed to be spent on this phase in comparison to the time taken in ISBIT-3.

The exact list of conformance tests performed in the first two phases of conformance testing and the results obtained by the three new products (the nine previously approved products had all been proven conformant during ISBIT-3) are shown in

Table 2 and its associated reference notes. It is apparent that the three new products, G, H and J passed all of the conformance tests. Therefore these three new products were declared conformant and allowed to proceed to the full performance based interoperability test.

Table 2. Conformance Testing Results

Product Tests	Reference Notes (see below)	G	H	J
ISBITAPI Initial Tests				
CaptureInit				
	1	Passed	Passed	Passed
CaptureEnd				
	2	Passed	Passed	Passed
Capture for Enrol				
	3	Passed	Passed	Passed
	4	Passed	Passed	Passed
	5	Passed	Passed	Passed
Capture for Verify				
	6	Passed	Passed	Passed
	7	Passed	Passed	Passed
	8	Passed	Passed	Passed
Enrol				
	9	Passed	Passed	Passed
	10	Passed	Passed	Passed
	11	Passed	Passed	Passed
	12	Passed	Passed	Passed
	13	Passed	Passed	Passed
	14	Passed	Passed	Passed
	15	Passed	Passed	Passed
	16	Passed	Passed	Passed
VerifyProcess				
	17	Passed	Passed	Passed
VerifyMatch				
	18	Passed	Passed	Passed
	19	Passed	Passed	Passed
	20	Passed	Passed	Passed
	21	Passed	Passed	Passed
	22	Passed	Passed	Passed
	23	Passed	Passed	Passed
	24	Passed	Passed	Passed
	25	Passed	Passed	Passed
	26	Passed	Passed	Passed
	27	Passed	Passed	Passed
	28	Passed	Passed	Passed
	BIR Conformance Verification			
	29	Passed	Passed	Passed
	30	Passed	Passed	Passed
	31	Passed	Passed	Passed
	32	Passed	Passed	Passed
	33	Passed	Passed	Passed
	34	Passed	Passed	Passed
	35	Passed	Passed	Passed
	36	Passed	Passed	Passed

Table 2 Reference Notes

1	Initialization of the sensor successful with zero (0) returned
2	Capture session is shutdown successfully (or no shutdown is required) and zero (0) is returned
3	GUI Prompts for correct finger when capturing finger image for Enrolment
4	Capture completes successfully with zero (0) returned and valid bitmap
5	User cancellation of capture completes successfully with negative two (-2) returned and valid bitmap
6	GUI Prompts for correct finger when capturing finger image for verification
7	Capture completes successfully with zero (0) returned and valid bitmap
8	User cancellation of capture completes successfully with negative two (-2) returned and valid bitmap
9	Enrolment returns zero (0) if 6 valid images provided (I,I,I,I,I,I)
10	Enrolment returns minus one (-1) using 3 Null for Primary Finger and 3 valid for Secondary Finger (N,N,N,I,I,I)
11	Enrolment returns zero (0) using 2 Null for Primary Finger and 3 valid for Secondary Finger (I,N,N,I,I,I)
12	Enrolment returns zero (0) using 1 Null for Primary Finger and 3 valid for Secondary Finger (I,I,N,I,I,I)
13	Enrolment returns minus two (-2) using 3 valid for Primary Finger and 3 Null for Secondary Finger (I,I,I,N,N,N)
14	Enrolment returns zero (0) using 3 valid for Primary Finger and 2 Null for Secondary Finger (I,I,I,I,N,N)
15	Enrolment returns zero (0) using 3 valid for Primary Finger and 1 Null for Secondary Finger (I,I,I,I,I,N)
16	Enrolment returns minus three (-3) if 6 Null images provided (N,N,N,N,N,N)
17	Successful processing of an input image into an intermediate template returns with zero (0)
18	Successful genuine match of primary finger intermediate template with BDIR from Test 9
19	Successful genuine match of secondary finger intermediate template with BDIR from Test 9 with UseSecondary (See Annex A, Section A.2.8)
20	Failed genuine match of primary finger intermediate template with BDIR from Test 9 with UseSecondary
21	Failed imposter match of primary finger intermediate template with BDIR from Test 9
22	Failed imposter match of secondary finger intermediate template with BDIR from Test 9 with UseSecondary
23	Successful genuine match of primary finger intermediate template with BDIR from Test 13
24	Failed genuine match of primary finger intermediate template with BDIR from Test 13 with UseSecondary
25	Failed imposter match of primary finger intermediate template with BDIR from Test 13
26	Failed imposter match of primary finger intermediate template with BDIR from Test 13 with UseSecondary
27	Failed genuine match of primary finger intermediate template with BDIR from Test 16
28	Failed genuine match of secondary finger intermediate template with BDIR from Test 16 with UseSecondary
29	BDIR returned from Test 9 meets all conformance assertions
30	BDIR returned from Test 10 meets all conformance assertions
31	BDIR returned from Test 11 meets all conformance assertions
32	BDIR returned from Test 12 meets all conformance assertions
33	BDIR returned from Test 13 meets all conformance assertions
34	BDIR returned from Test 14 meets all conformance assertions
35	BDIR returned from Test 15 meets all conformance assertions
36	BDIR returned from Test 16 meets all conformance assertions

4 Performance Based Interoperability Results and Data Analysis

This section presents a summary of the interoperability and performance test results. It is critical for an understanding of how the final conclusions were achieved.

4.1 Introduction and Important Notes

Before beginning to consider the results themselves, it is important to recognize that this test, like any other biometric conformance, performance and interoperability test, is subject to a number of caveats. The foremost of these is the limitation in the size of the test crew, the group of test subjects that participate in the test.

The ILO tests products consisting of fingerprint sensors combined with enrolment and matching algorithms. It is therefore necessary to use a live capture test any time a new sensor is to be introduced, and since the interoperable performance of the product using that sensor must be compared with that of all other previously approved products, an ideal test that does the comparison with the same test crew and in the same environment will perform live fingerprint capture for all products every time a new sensor is to be added. This is why the ILO ISBIT tests only occur periodically when there are a sufficient number of new products to justify the cost of a live capture data acquisition across multiple products. A new algorithm can be introduced that uses images captured by a previously tested sensor and this does not require that an entire new database of fingerprint images be captured, but it requires that the new product continue to use the exact same sensor and image capture software as the previously approved product used for the image capture. This occurred in ISBIT-3 when one of the previously approved vendors wished to update their algorithm and it allowed their old algorithm and their new algorithm to be tested as if they were two separate products without having to collect a second set of fingerprint images. It also occurred in ISBIT-4, when modified algorithms were substituted for old algorithms and the entire interoperability matrix was recalculated, as described in Annex D. In general, however, live capture is required and every member of the test crew will have to enrol two fingers and verify multiple fingers across all of the products being tested on at least two visits, separated by two or three weeks. This can be very time consuming and exhausting for the test subjects, so that they require a significant payment to convince them to participate and close supervision to ensure that they do not become fatigued and start to make mistakes when following the instructions of the biometric products. Another problem arises from the fact that many potential test crew are uncomfortable with providing fingerprints for a database, no matter what assurances they are given on the use of that database. This makes recruitment of the test crew more difficult and time consuming and also results in a percentage of test crew that revoke their consent at different stages in the process, making the management of the test more difficult. All of this means that there are practical limits to the size of the test crew, based on the time and financial constraints of the test.

Another fundamental limitation is that the ILO performance metrics are based on two finger transactions, as described in ILO SID-0002. This is highly positive, in that the test results are more likely to predict real world performance if they are based on the same transactions that will be used in the real world. It is a limitation, however, because it means that a full enrolment process must be used, potentially resulting in up to 30 single finger placements (3 placements of up to 10 fingers) before a failure to enrol can be declared and a complete verification transaction of 6 finger placements (3 finger placements for each of 2 fingers) must be recorded in order to properly compute the performance of a two finger transaction. Once again this impacts the number of independent transactions that can be recorded in a reasonable period of time.

In order to maximize the amount of available data for enrolment and verification with approximately two weeks between the enrolment visit and the verification visit, a technique was used that was first developed for ISBIT-3. During the offline matching, the verification images from Visit 2 were matched against the enrolment BDIRs generated during Visit 1 and the verification images from Visit 1 were matched against the enrolment BDIRs generated during Visit 2. This meant there was always the same amount of time between enrolment and verification, but in one half of the data enrolment occurred prior to verification and vice versa in the other half of the data. This may result in subtle biases between the two halves of the data. As an example, it is quite conceivable that the second set of enrolment records might show better performance, since the test subject might be more familiar with the products during the second visit. Results from ISBIT-3, however, did show that the differences in interoperable performance calculated using these two halves of the match data were less than the statistical uncertainty in the measurements themselves and therefore this is only a minor caveat about the validity of the ISBIT-4 results.

A separate issue from the limits on the size of the data set used for testing is the fact that it is very difficult to compare biometric performance data from one test to another. Aside from standard variations in the results that are within the uncertainty calculated using normal statistical methods, there are a number of factors peculiar to biometric testing that may cause systematic biases between one test and another. The composition of the test crew is one such factor, since a small number of people with especially difficult fingers can have a significant effect on performance, especially if they have multiple failures to enrol. The ISBIT-4 test crew was 68% composed of people from the ISBIT-3 test crew and therefore this issue was mitigated, but not completely eliminated. Unfortunately 100% overlap between test crews on experiments separated by years is very difficult to achieve. One option was to ignore the data from those people who were not part of the ISBIT-3 crew, but this would have reduced the test crew size to 125, which was considered to be too small. Another related factor is the difference in environmental conditions between one test and another. Even though the test took place in an office environment, the external seasonal environment may have a significant effect on the overall condition of the test subjects' fingerprints. ISBIT-3, for instance took place in the middle of Winter and many test subjects suffered from problems with dry skin. ISBIT-4 took place in the Autumn, when the outside environment was quite damp and external temperatures were still above freezing. Additional systematic errors may have been caused by differences in the test administrators and their interaction with the test subjects or by the slight differences in the test setup such as the fact that all biometric products were present at a single station and so there was no need for test subjects to change test stations as they did during the data capture in ISBIT-1 and ISBIT-3.

Such systematic biases are present in every biometric test that captures new data from live test subjects and therefore a method of normalizing for them when deciding on an appropriate group of interoperable products is required. That is why the interoperable performance criterion for inclusion in the list of ILO products that was described in Section 1.4 above was used.

Another important point to note is that all of the performance numbers quoted in this report are based on transactional error rates, including transactional failure to enrol (FTE) and transactional failure to acquire (FTA). It is important to define what this means. An enrolment transaction started with the test subject providing three presentations of their right index finger. The images from all of the three presentations (except those returned as "failure to acquire" by the biometric product's "Capture" function) were submitted to the product's "Enrol" function in the primary finger image positions with three nulls in the secondary finger image positions. If the "Enrol" function reported a "failure to enrol" then the next finger in the list defined in ILO SID-0002 was selected and three presentations were acquired from it. These were then submitted to the product's "Enrol" function, with a failure resulting in the next finger in the list being selected. At any point at which the "Enrol" function declared success, it would return a BDIR with the primary finger enrolled and the secondary finger as an "Unenrolled Finger".

The set of three presentations from the next finger would then be submitted to the product's "Enrol" function in the secondary finger image positions with the three images (or a combination of images and nulls if some presentations had a "failure to acquire") from the primary finger that was just enrolled successfully in the primary finger image positions. This would continue until the "Enrol" function returned a BDIR that had two fingers successfully enrolled or until the test subject had no more fingers left to try to enrol. Since a BDIR with only a primary finger enrolled was considered acceptable for use in the offline verification portion of the test, a true transactional failure to enrol required that a total of 3 presentations of each of ten different fingers had to result in no successfully acquired images that were able to be processed for enrolment. Similarly, a transactional failure to acquire on verification meant that a total of three presentations of each of the primary and secondary fingers (or primary only if the enrolment BDIR contained only a single finger) had to result in a "failure to acquire".

Overall, this meant that there were only 3 transactional failure to enrol errors in the entire test out of a total of 4536 enrolment transactions. There were also a total of 148 failure to acquire errors out of a total of 20,251,227 two finger simulated transactions.

The false match rate (FMR) and false non-match rate (FNMR) were also calculated using transactional scores. A false match occurred when an imposter achieved a match score above the match threshold on at least one of the six (or three) finger presentations during a verification transaction. Note that the choice of six or three transactions did not depend on the imposter, but on whether the genuine that they were attempting to match against was able to enrol two fingers or only one. A false non-match occurred when a genuine did not achieve a match score equal to or greater than the match threshold on any of the six (or three) finger presentations. This could be caused by any mix of low match scores and single presentation failures to acquire. Given these transactional definitions of FMR, FNMR, FTA and FTE, the transactional generalized false accept rate (GFAR) and generalized false reject rate (GFRR) that are relevant to the ILO interoperable performance threshold can be calculated using the standard formulas:

$$GFAR = FMR * (1-FTA) * (1-FTE)$$

GFRR = FTE + (1 – FTE) * FTA + (1 – FTE) * (1 – FTA) * FNMR

The final point to note is related to confidence intervals, and is particularly a problem for performance interoperability testing such as that carried out in ISBIT-4. Most estimates of confidence require large numbers of independent samples. Given the limitations in the size of the test crew already noted above, the test data contains multiple enrolment and verification attempts by each test subject and the underlying distribution of those is probably different than it would have been if every enrolment and verification attempt had been performed by a different test subject. Of course the underlying distribution in either case is always difficult to determine, especially in biometric testing. Furthermore, extrapolating the results of the test to real world performance with deployed systems requires some estimate of how well the test environment models expected operating conditions for a deployed system, how well the test crew models the demographics of the expected population of users and ideally requires verification visits at multiple dates after the enrolment visit up to the ten year lifespan of an SID. Unfortunately, this data is simply not available and therefore there may be unknown systematic differences between the performance measured in this test and that which will be observed in real world deployments of the ILO SID. Despite this, ISBIT-4 is a reasonable attempt to model the enrolment and verification transactions defined by ILO with a variety of fingerprint technologies using a test crew with a cross-section of ethnicity and job category. Given all of the unknowns, the measurements are assumed to follow a normal distribution and the uncertainty in the measurements has been calculated on this basis.

4.2 DET Curves

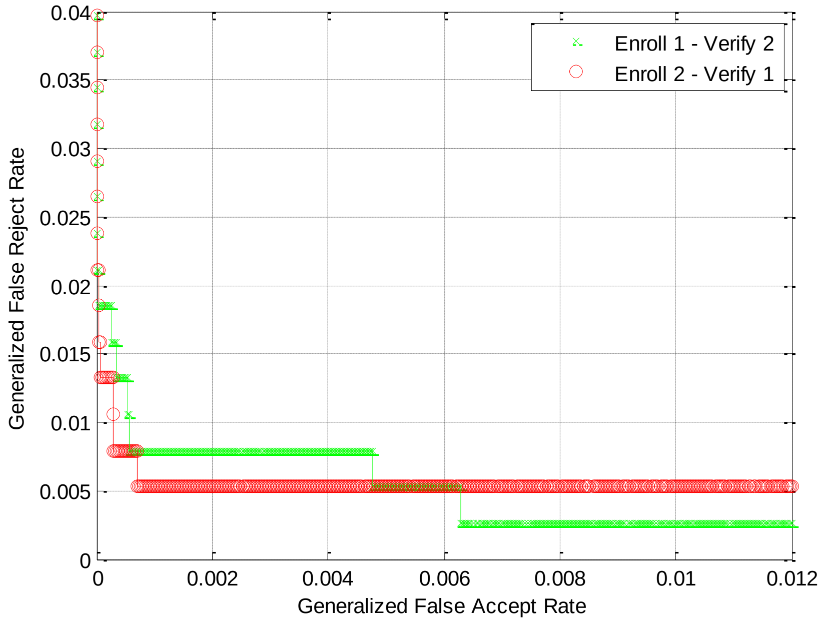
For each of the possible combinations of enrolment product and verification product, there was a natural breakdown in the data to allow two distinct Detection Error Trade-off Curves to be computed. These used the following subsets of the data:

1. Enrol during Visit 1, Verify during Visit 2 (E1V2)
2. Enrol during Visit 2, Verify during Visit 1 (E2V1)

Although the final performance interoperability results are based on the combined performance of these two datasets, it is worthwhile to compare the results with enrolment approximately two weeks ahead of verification second to the results with verification data acquired approximately two weeks prior to the enrolment data. This allows an examination of whether or not the time difference between enrolment and verification being negative or positive really did make a difference in the results.

The following figure shows a DET curve for enrolment on Product I and verification on Product I using the metric of transactional GFRR on the y-axis and transactional GFAR on the x-axis.

Figure 6. Enrol H Verify L – GFRR Versus GFAR



The GFRR values are quantized because of the limited number of genuine transactions (approximately 378 complete two finger transactions in most cases) in each of these two data sub-sets for a specific enrol product – verify product combination. The GFAR values are not significantly quantized, as the exhaustive offline testing supported a very large number of imposter transactions (typically around 70,000 per enrol product – verify product combination for each of E1V2 and E2V1). Subject to the limits of the quantization, it appears that there is no significant difference between the two data sets, suggesting that the separation between enrolment and verification can be either backwards or forwards in time.

In Annex E, there are complete sets of DET figures for all of the combinations of enrol and verify products.

4.3 Performance Interoperability Matrices

The performance interoperability matrices show the GFRR at a 1% GFAR (known as G^1) for all possible combinations of enrol and verify products. The ILO criterion for inclusion in the approved products list is that the group of products selected (which must include the previously qualified products) shall have an interoperability matrix with a mean that is either less than 1% or less than the mean of the interoperability matrix containing only the previously qualified products and also with a maximum that is less than the maximum of the interoperability matrix containing only the previously qualified products. In more precise terms, the interoperable performance criterion can be defined as follows:

- Let G^{MEAN} be the mean of the G^1 values for all nine previously approved products
- Let G^{MAX} be the maximum of the G^1 values for all nine previously approved products

- Let G_2^{MEAN} be the mean of the G^1 values of the nine previously approved products and those new products being considered for a potential new list of approved products
- Let G_2^{MAX} be the maximum of the G^1 values of the nine previously approved products and those new products being considered for a potential new list of approved products
- The following two conditions must then be satisfied for the new list of products to be approved:

1. G_2^{MEAN} is less than or equal to G^{MEAN} **OR** G_2^{MEAN} is less than 1%
2. G_2^{MAX} is less than or equal to G^{MAX}

Note that as in previous ISBIT tests, all GFRR values are calculated based on an operating point producing a 1% GFAR on the DET curve associated with that particular enrol product, verify product combination. Therefore the operating thresholds are not constant for each column.

The first important interoperability matrix is the one that shows the results only for the nine previously qualified products, which in ISBIT-4, due to the random renaming of each product, are denoted as A, B, C, D, E, F, I, K and L. Note that all GFRR results are shown as percentages and that the native performance of each product (that is the performance when the same product is used for enrolment and verification) is in **bold** text and lies on the diagonal of the matrix. The elements of the matrix are shaded more darkly as the GFRR value increases, so it is easy to get an overall visual sense of relative interoperable performance. At the end of each row and column, the mean value is provided. This is indicative of the total interoperable performance of that product for enrolment or verification and the rankings are based exclusively on the mean. All values except the overall mean of the entire interoperability matrix have been rounded to a single decimal place, as this is appropriate given the uncertainty in the measurements, but the ranking are based on the mean values before rounding.

Table 3. Previously Approved Products Interoperability Matrix, GFRR at GFAR = 1%

Enrol Product \ Verify Product	A	B	C	D	E	F	I	K	L	MEAN	RANK
A	**1.6**	0.0	0.7	1.1	0.5	0.0	0.1	1.7	0.5	0.7	6
B	1.6	**0.0**	0.7	0.5	0.1	0.1	0.1	1.2	0.5	0.5	4
C	0.9	0.5	**0.7**	0.8	0.7	0.5	0.5	0.7	0.7	0.7	5
D	1.5	0.4	1.1	**0.3**	0.7	0.4	0.8	1.9	0.8	0.9	7
E	1.1	0.1	0.8	0.0	**0.0**	0.0	0.0	0.5	0.3	0.3	1
F	0.9	0.5	0.8	0.4	0.3	**0.0**	0.1	1.1	0.5	0.5	3
I	1.2	0.1	0.8	0.3	0.5	0.1	**0.0**	0.7	0.3	0.4	2
K	2.8	1.1	0.9	1.2	0.9	0.3	0.4	**0.9**	1.3	1.1	8
L	2.2	0.7	1.3	1.5	0.9	0.7	1.1	1.9	**0.8**	1.2	9
MEAN	1.5	0.4	0.8	0.7	0.5	0.2	0.4	1.2	0.6	**0.70**	
RANK	9	3	7	6	4	1	2	8	5		

In ISBIT-3 these nine products obtained a mean GFRR of 0.92%, whereas in this test they obtained a mean G^1 of G^{MEAN} = 0.70%. Similarly, in ISBIT-3 these nine products obtained a maximum G^1 of 2.4%, whereas in this test they obtained a maximum of G^{MAX} = 2.8%. The increase in G^{MAX} is well within the statistical uncertainty of a single entry in the interoperability matrix, but the decrease in G^{MEAN} is more than the statistical uncertainty of 0.06% calculated in the ISBIT-3 test report. This may, however, be due to the more humid conditions present in this test in comparison to ISBIT-3, since the total number of transactional failure to enrol and failure to acquire errors observed in this test was less than in the previous test. The native performance found on the diagonal ranges from zero recorded errors to GFRR = 1.6%. In ISBIT-3, the values ranged from 0.1% to 1.5%.

Next consider the interoperability matrix corresponding to all twelve products tested. This includes the nine previously approved products and the three new products designated as G, H and J.

Table 4. All Products Interoperability Matrix, GFRR at GFAR = 1%

Enrol Product \ Verify Product	A	B	C	D	E	F	G	H	I	J	K	L	MEAN	RANK
A	**1.6**	0.0	0.7	1.1	0.5	0.0	1.7	0.5	0.1	0.5	1.7	0.5	0.7	7
B	1.6	**0.0**	0.7	0.5	0.1	0.1	1.3	0.3	0.1	0.5	1.2	0.5	0.6	4
C	0.9	0.5	**0.7**	0.8	0.7	0.5	0.5	0.5	0.5	0.5	0.7	0.7	0.6	5
D	1.5	0.4	1.1	**0.3**	0.7	0.4	0.8	1.2	0.8	0.9	1.9	0.8	0.9	9
E	1.1	0.1	0.8	0.0	**0.0**	0.0	0.7	0.0	0.0	0.7	0.5	0.3	0.3	1
F	0.9	0.5	0.8	0.4	0.3	**0.0**	1.5	0.0	0.1	0.7	1.1	0.5	0.6	3
G	2.1	1.3	1.6	1.3	1.3	1.3	**1.6**	1.3	1.3	1.4	2.5	1.1	1.5	12
H	2.1	0.7	1.1	0.9	0.4	0.0	0.9	**0.0**	0.3	0.9	1.3	0.4	0.7	6
I	1.2	0.1	0.8	0.3	0.5	0.1	0.9	0.0	**0.0**	0.7	0.7	0.3	0.5	2
J	1.5	0.5	0.7	0.7	0.7	0.5	1.2	0.5	0.5	**0.5**	1.2	0.7	0.8	8
K	2.8	1.1	0.9	1.2	0.9	0.3	2.0	1.2	0.4	1.1	**0.9**	1.3	1.2	10
L	2.2	0.7	1.3	1.5	0.9	0.7	1.6	0.8	1.1	1.1	1.9	**0.8**	1.2	11
MEAN	1.6	0.5	0.9	0.7	0.6	0.3	1.2	0.5	0.4	0.8	1.3	0.7	**0.80**	
RANK	12	3	9	7	5	1	10	4	2	8	11	6		

The most important conclusion to draw from this Table is that G_2^{MEAN} = 0.80% and G_2^{MAX} = 2.8%. Native performance also varies from zero recorded errors to 1.6%, exactly as it did before. Therefore, it would appear that all three new products should be added to the approved products list, since this satisfies both of the criteria for approval in that:

1. $G_2^{MEAN} < 1\%$
2. $G_2^{MAX} = G^{MAX}$

There is one slight issue with product G. Although it successfully passed all of the conformance tests prior to the start of the data acquisition phase, there were a total of two templates that it produced during data acquisition which turned out to be non-conformant. These templates were non-conformant in a very specific way. They had two minutiae with different angles in the same x and y position in the template. This is not explicitly forbidden by the ISO 19794-2 fingerprint minutiae standard, but it is considered to be an implicit requirement of the standard that all minutiae be defined in accordance with the definitions in the standard and therefore it should not be possible to have two different minutiae at the same location. When these two templates were used for matching, however, they did not result in any errors and therefore this particular conformance issue did not seem to be a problem for interoperability. The results in Table 4 therefore include the two non-conformant templates that were produced by Product G. Ideally, no approved product should produce any type of non-conformant template and therefore the approval of Product G was uncertain. The supplier of Product G agreed to provide an update to the algorithm that would fix the problem with multiple minutiae at the same position and let this be tested together with the two modified algorithms that had been provided from suppliers of previously approved products. The results of this test are described in Annex D,

Further analysis Table 4 leads to the conclusion that some products (such as E) are better at generating interoperable biometric information records (BDIRs) and some products (such as F) are better at matching against BDIRs generated by other products. Of course in this matrix the FTE and FTA values, which are potentially quite dependent on the sensors used in each product, are built in to the results.

There is also some value in considering a performance interoperability matrix defined by FNMR at a 1% FMR. This gives results that can be used to evaluate the algorithms if every image returned from their corresponding sensors was of sufficient quality. It is not realistic for predicting real world performance, but does provide interesting insights.

Table 5. All Products Interoperability Matrix, FNMR at FMR = 1%

	A	B	C	D	E	F	G	H	I	J	K	L	MEAN	RANK
A	**1.2**	0.0	0.1	1.1	0.5	0.0	1.5	0.5	0.1	0.0	1.7	0.5	0.6	8
B	1.2	**0.0**	0.1	0.5	0.0	0.1	0.9	0.3	0.1	0.0	1.2	0.5	0.4	6
C	0.5	0.3	**0.0**	0.5	0.0	0.3	0.3	0.3	0.3	0.1	0.4	0.4	0.3	3
D	1.2	0.4	0.8	**0.3**	0.5	0.4	0.4	1.2	0.8	0.3	1.9	0.8	0.7	9
E	0.8	0.1	0.3	0.0	**0.0**	0.0	0.5	0.0	0.0	0.1	0.5	0.3	0.2	2
F	0.5	0.5	0.3	0.4	0.3	**0.0**	1.2	0.0	0.1	0.0	1.1	0.5	0.4	5
G	1.7	1.3	1.1	1.3	1.3	1.3	**1.3**	1.3	1.3	0.8	2.5	1.1	1.4	12
H	1.7	0.7	0.5	0.9	0.4	0.0	0.5	**0.0**	0.3	0.3	1.3	0.4	0.6	7
I	0.8	0.1	0.3	0.3	0.5	0.1	0.7	0.0	**0.0**	0.0	0.7	0.3	0.3	4
J	0.8	0.0	0.0	0.1	0.0	0.0	0.7	0.0	0.0	**0.0**	0.7	0.1	0.2	1
K	2.4	1.1	0.3	1.2	0.9	0.3	1.7	1.2	0.4	0.4	**0.9**	1.3	1.0	10
L	1.9	0.7	0.8	1.5	0.9	0.7	1.3	0.8	1.1	0.4	1.9	**0.8**	1.0	11
MEAN	1.2	0.4	0.4	0.7	0.5	0.3	0.9	0.5	0.4	0.2	1.2	0.6	**0.60**	
RANK	12	5	4	9	6	2	10	7	3	1	11	8		

Verify Product (columns); Enrol Product (rows)

As expected, the overall performance is improved, with the mean of the entire performance interoperability matrix dropping to 0.60%. There are also many zeroes in the matrix, indicating that some products could interoperate without any recorded false non-match errors in the ISBIT-3 test. Perhaps most interesting is the case of the new product J. In Table 4, this product was slightly below average, with a ranking of 8 as both a verify and enrol product. In Table 5 it moved to the top ranking as both an enrol and verify product. This probably means that is has an excellent template generation and matching algorithm but that either the sensor or the data acquisition software are poorly designed so that it rejects many fingerprints that other products would attempt to use. The new product H, on the other hand, has virtually no difference between its performance in Table 4 and Table 5. It scores above average when FTA and FTE are considered in Table 4, but ranks about average or even slightly below when no such errors are considered in Table 5. This probably means that it has an excellent sensor and data acquisition software which always captures a reasonable image, but that its template generation and matching software is only of average performance. All of this shows the importance of measuring sensor effects. Unless real data acquisition take place so that FTE and FTA can be properly measured, the ranking of the products may be quite misleading.

4.4 Confidence Intervals

It is very difficult to provide a reasonable estimate for the confidence interval on the results in a performance interoperability test such as this. There are numerous reasons, some of them mentioned in Section 4.1, why there is no direct way of estimating the underlying distribution of scores and thus the uncertainty in each measurement. There are some simple methods that may help to give some estimate of the confidence, however, and a brief analysis follows, based on the methods outlined in the standard ISO 19795-4:2008 Biometric Performance Testing and Reporting Part 4 – Interoperability Performance Testing.

There are approximately 756 genuine transactions representing approximately 4536 single finger matches used to compute the GFRR value in each position in the performance interoperability matrix. The average of these values is 0.80%. Being conservative, a one sided 95%, confidence interval based on the assumption of 756 independent measurements and a normal distribution can be used to find the upper limit for the performance interoperability numbers. If we make a rough estimate of a chance of non-match for a genuine in each position in the interoperability matrix being p = 0.01 (1%), then we get:

$$Z_{95\%,\ 1\ \text{sided}} = 1.645\ (p * (1-p) / 756)^{0.5} = 0.006 = 0.6\%$$

This means that there is only a 5% chance that each of the entries in the matrix will be more than 0.6% larger than it is now. Actually, this is oversimplifying, because the value of the confidence interval depends on the currently measured value, and these vary across the interoperability matrix, but 0.6% is a reasonable rule of thumb.

Of more relevance is the possibility that the entire performance interoperability matrix will be shifted, resulting in a mean that changes significantly from its current value. Since there are 144 entries in the performance interoperability matrix, we can run the same calculation for the mean, using 144 * 756 = 108864 independent measurements and using the exact value of p = 0.0080. This gives a 95%, one sided confidence interval of 0.04%, suggesting that there is only a 5% chance that the mean GFRR would be larger than 0.84%. This gives good confidence that the ILO requirement of mean interoperable performance of GFRR less than or equal to 1% at a GFAR of 1% will continue to be met if this experiment is repeated, provided of course that systematic changes such as environmental conditions, demographics, the type of guidance given by the administrators to the test crew during enrolment and verification, etc. do not substantially affect the result.

One problem with the analysis above is, of course, that the measurements are not truly independent. Only 84,806 fingerprint images were captured and the simulated transactions used to compute GFRR at 1% GFAR necessarily involve some reuse of the images. Another option for exploring the confidence in the data is to break it back into the two separate data sets representing enrolment during the first visit and verification during the second visit (E1V2) versus enrolment during the second visit and verification during the first visit (E2V1). These two sets are completely independent and although the natural variability in GFRR at 1% GFAR in each of these data sets will be larger than for the combined data set (approximately 0.06% for a 95%, one sided confidence interval since each subset uses half of the total genuine match transactions), the difference between them will help to evaluate the confidence in the results.

The two tables below show the full performance interoperability matrix based on GFRR at a GFAR of 1% for E1V2 and then for E2V1.

Table 6. All Products Interoperability Matrix for E1V2 Only, GFRR at GFAR = 1%

Enrol Product \ Verify Product	A	B	C	D	E	F	G	H	I	J	K	L	MEAN	RANK
A	**1.6**	0.0	0.5	0.8	0.5	0.0	1.8	0.3	0.3	0.5	2.4	0.8	0.8	7
B	1.3	**0.0**	0.5	0.3	0.0	0.0	0.8	0.0	0.3	0.5	0.8	0.3	0.4	3
C	0.8	0.5	**0.5**	1.1	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6	5
D	1.6	0.5	1.1	**0.5**	0.5	0.3	0.5	1.1	0.5	1.1	1.9	1.1	0.9	8
E	0.8	0.0	0.8	0.0	**0.0**	0.0	0.8	0.0	0.0	0.8	0.8	0.5	0.4	1
F	1.3	0.5	0.5	0.5	0.3	**0.0**	1.6	0.0	0.3	0.5	1.1	0.5	0.6	4
G	1.3	1.1	1.1	1.1	1.1	1.1	**1.6**	1.1	1.1	1.1	1.9	0.8	1.2	11
H	2.6	1.1	1.1	1.3	0.8	0.0	1.1	**0.0**	0.5	1.1	2.1	0.3	1.0	9
I	0.8	0.0	0.5	0.3	0.5	0.0	0.8	0.0	**0.0**	0.5	0.8	0.3	0.4	2
J	1.3	0.5	0.5	0.5	0.5	0.5	1.6	0.5	0.5	**0.5**	1.1	0.8	0.7	6
K	3.2	1.1	0.7	1.3	1.3	0.0	1.3	0.8	0.3	1.1	**1.1**	0.8	1.1	10
L	2.9	0.3	1.3	1.9	1.1	0.5	2.1	0.5	1.1	1.1	2.4	**0.5**	1.3	12
MEAN	1.6	0.5	0.8	0.8	0.6	0.2	1.2	0.4	0.4	0.8	1.4	0.6	**0.77**	
RANK	12	4	7	9	5	1	10	2	3	8	11	6		

Table 7. All Products Interoperability Matrix for E2V1 Only, GFRR at GFAR = 1%

	A	B	C	D	E	F	G	H	I	J	K	L	MEAN	RANK	Enrol Product
A	**1.6**	0.0	0.8	1.3	0.5	0.0	1.6	0.8	0.0	0.5	1.1	0.3	0.7	6	
B	1.9	**0.0**	0.8	0.8	0.3	0.3	1.9	0.5	0.0	0.5	1.6	0.8	0.8	8	
C	1.1	0.5	**0.8**	0.5	0.8	0.5	0.5	0.5	0.5	0.5	0.8	0.8	0.7	5	
D	1.3	0.3	1.1	**0.0**	0.8	0.5	1.1	1.3	1.1	0.8	1.9	0.5	0.9	9	
E	1.3	0.3	0.8	0.0	**0.0**	0.0	0.5	0.0	0.0	0.5	0.3	0.0	0.3	1	
F	0.5	0.5	1.1	0.3	0.3	**0.0**	1.3	0.0	0.0	0.8	1.1	0.5	0.5	3	
G	2.9	1.6	2.1	1.6	1.6	1.6	**1.6**	1.6	1.5	1.8	3.2	1.3	1.9	12	
H	1.6	0.3	1.1	0.5	0.0	0.0	0.8	**0.0**	0.0	0.8	0.5	0.5	0.5	2	
I	1.6	0.3	1.1	0.3	0.5	0.3	1.1	0.0	**0.0**	0.8	0.5	0.3	0.6	4	
J	1.6	0.5	0.8	0.8	0.8	0.5	0.8	0.5	0.5	**0.5**	1.3	0.5	0.8	7	
K	2.4	1.1	1.1	1.1	0.5	0.5	2.6	1.6	0.5	1.1	**0.8**	1.9	1.3	11	
L	1.6	1.1	1.3	1.1	0.8	0.8	1.1	1.1	1.1	1.1	1.3	**1.1**	1.1	10	
MEAN	1.6	0.5	1.1	0.7	0.6	0.4	1.2	0.7	0.4	0.8	1.2	0.7	**0.83**		
RANK	12	3	9	6	4	1	11	5	2	8	10	7			
	Verify Product														

Although there are individual elements in the two interoperability matrices that are different, the means of the two matrices are 0.77% and 0.83% respectively. This means that the difference between them is within the computed 95% uncertainty of 0.06% and suggests that this computation of the uncertainty in the experiment may be reasonable.

5 Interoperable Product Combinations

ISBIT-4 was designed to determine whether or not the ILO approved products list could be expanded beyond the existing nine products to include one or more of the three additional products (G, H and J) being tested. Any such products had to satisfy the conformance, performance and interoperability requirements of the ILO. Since all products evaluated were conformant (although one had to be retested with a modified algorithm as described in Annex D to prove this) and since the mean of the performance interoperability matrix containing the nine previously approved products and the three new products was 0.80%, it was clear that all twelve products met the requirements and should be recommended for inclusion on the approved products list. It is interesting, however, to see what the interoperable performance would be if various subsets of the twelve products were used. The table below lists the mean and maximum GFRR of the performance interoperability matrix for the best performing sub-group of the twelve products at various sizes of sub-group.

Table 8. Summary of Best Performing Product Combinations

Results Constrained by Interoperability with Current ILO Approved Products			
Size of Sub-Group	**Product Identifiers**	**Mean GFRR**	**Max GFRR**
9	A, B, C, D, E, F, I, K, L	0.70%	2.77%
10	A, B, C, D, E, F, H, I, K, L	0.68%	2.77%
11	A, B, C, D, E, F, H, I, J, K, L	0.69%	2.77%

This shows that with the constraint that the existing products must be included, there is essentially no difference in performance when Product H and Product I are added to the approved products list. The addition of Product G moves the mean GFRR to 0.80%, as shown in Table 4 above, which means that it does decrease performance, albeit by a small enough margin that it still qualifies as an approved product. For further comparison purposes, the table below provides the same analysis if there was no requirement to include the nine previously approved products. It shows that the new Product H is actually one of the best performing interoperable products. This is important, since in Table 4, Product H is in the top half of all products, but certainly doesn't appear to be one of the two best products. Most of this appears to be caused by its poor performance with the previously approved products A and K, however, and without the requirement to support these products, the overall interoperability matrix is improved substantially and product H turns out to be one of the two best performing products.

Table 9. Summary of Best Performing Product Combinations without Constraints

Results Without Constraints of Current ILO Approved Products			
Size of Sub-Group	**Product Identifiers**	**Mean GFRR**	**Max GFRR**
2	F, H	0.00%	0.00%
3	F, H, I	0.06%	0.26%
4	E, F, H, I	0.11%	0.53%
5	B, E, F, H, I	0.15%	0.66%
6	B, D, E, F, H, I	0.27%	1.19%
7	B, D, E, F, H, I, J	0.37%	1.19%
8	B, C, D, E, F, H, I, J	0.45%	1.19%
9	B, C, D, E, F, H, I, J, L	0.51%	1.46%
10	A, B, C, D, E, F, H, I, J, L	0.60%	2.24%
11	A, B, C, D, E, F, H, I, J, K, L	0.69%	2.77%

6 Conclusions and Recommendations

Based on the results of the ISBIT-4 test, the following conclusions can be made:

1. All three of the new products (G, H and J) meet the ILO requirements for conformance, performance and interoperability and should be added to the approved products list, although product G is now defined to use the modified algorithm which achieves the results shown in Annex D.

2. The modified algorithm for Product F achieved the performance described in Annex D and satisfies the ILO requirements for conformance, performance and interoperability. This modified version is therefore included in the recommended list of approved products provided to ILO as an attachment to this report. The modified algorithm for Product D did not meet the requirements and the so original version of Product D should continue to be used.

3. It is possible for biometric products to produce conformant BDIRs in almost all cases and only occasionally produce non-conformant data records. Sometimes a product thought to be conformant on the basis of one test, may be determined to be non-conformant under certain circumstances in a larger test, as happened with Product G when it moved from the conformance testing phase to the data acquisition phase in ISBIT-4. Therefore it is important to test every record produced in a deployed system to ensure it is conformant prior to encoding it in an identity document or database.

7 References

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

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

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

d. Biometric Testing Campaign Report (Addendum to Part 1)

e. 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)

f. ISO/IEC 19795-1:2006 Biometric performance testing and reporting -- Part 1: Principles and framework

g. ISO/IEC 19795-2:2007 Biometric performance testing and reporting -- Part 2: Testing methodologies for technology and scenario evaluation

h. ISO/IEC 19795-4:2008 Biometric performance testing and reporting -- Part 4: Interoperability performance testing

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. Some of the terms, however, have been given narrower definitions than in the reference standards due to their use in the specific context of testing conformance, performance and interoperability of biometric products to be used with ILO Seafarers' Identity Documents. Specific relevant terms are defined below:

administrator
person performing the testing or enrolment

attempt
submission of one (or a sequence of) biometric samples to the system

NOTE An attempt results in an enrolment template, a matching score (or scores), or possibly a failure to acquire.

Biometric Interchange Record (BDIR)
refers to a ILO SID-0002 conformant data record containing up to two fingerprint minutiae templates

test crew
set of test subjects gathered for an evaluation

detection error trade-off (DET) curve
modified ROC curve which plots error rates on both axes (false positives on the x-axis and false negatives on the y-axis)

enrolment
application in which the user is processed by a system in order to generate and store an enrolment template for that individual

enrolment attempt
the submission of three enrolment presentations of one finger on the part of a user for the purpose of enrolment in a biometric system

enrolment presentation
the submission of a single biometric characteristic (fingerprint) on the part of a user for the purpose of enrolment

enrolment transaction
sequence of up to 10 enrolment attempts (one per finger) on the part of a user resulting in an enrolment or a failure to enrol

experimenter

person responsible for defining, designing, and analyzing the test

failure to acquire rate (FTA)
proportion of verification transactions for which the system fails to capture or locate an image or signal of sufficient quality

failure to enrol rate (FTE)
proportion of the population for whom the system fails to complete the enrolment process

NOTE The observed failure to enrol rate is measured on test crew enrolments. The predicted/expected failure to enrol rate will apply to the entire target population.

false accept rate (FAR)
proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed

false match rate (FMR)
proportion of zero-effort impostor attempt samples falsely declared to match the compared non-self template

NOTE The measured/observed false match rate is distinct from the predicted/expected false match rate (the former may be used to estimate the latter).

false non-match rate (FNMR)
proportion of genuine attempt samples falsely declared not to match the template of the same characteristic from the same user supplying the sample

NOTE The measured/observed false non-match rate is distinct from the predicted/expected false non-match rate (the former may be used to estimate the latter).

false reject rate (FRR)
proportion of verification transactions with truthful claims of identity that are incorrectly denied

generalized false accept rate (GFAR)
metric for false accept rate that includes the effect of failures to enrol and failures to acquire. Specifically, GFAR = FMR * (1-FTA) * (1-FTE)

generalized false reject rate (GFRR)
metric for false reject rate that includes the effect of failures to enrol and failures to acquire. Specifically, GFRR = FTE + (1 – FTE) * FTA + (1 – FTE) * (1 – FTA) * FNMR

genuine attempt
single good-faith attempt by a user to match their own stored template

guidance
direction provided by an administrator to a test subject in the course of data capture for enrolment or verification

NOTE Guidance is separate from feedback provided by a biometric system or device in the course of data capture, such as audible or visual presentation queues.

habituation
the degree of familiarity a test subject has with a device

NOTE A test subject with substantial familiarity using a biometric device, such as that gained in the course of employment, is referred to as a habituated test subject.

impostor attempt
see *zero-effort impostor attempt*

intermediate template
biometric *sample* generated or processed to conform to a vendor's own closed unknown format

interoperability
measure expressing the verification performance associated with the use by vendor A of biometric data conforming to a standard interchange format generated by vendor B or vice versa

match attempt
the submission of three match presentations on the part of a user for matching in a biometric system

match presentation
the submission of a single biometric characteristic (fingerprint) on the part of a user for matching

match transaction
sequence of two match attempts (corresponding with two templates in a BDIR) on the part of a user simulated during offline testing resulting in a verification decision

NOTE If a BDIR only contains a single enrolled template, a match transaction will consist of a single match attempt.

native verification
a verification in which the claimed identity template was enrolled using the same biometric product as is used to verify the user

non-native verification
a verification in which the claimed identity template was enrolled using a different biometric product than is used to verify the user

observer
test staff member recording test data or monitoring the crew

offline testing
execution of enrolment and matching separately from data capture

NOTE 1 Collecting a database of samples for offline enrolment and calculation of matching scores allows greater control over which samples and attempts are to be used in any transaction.

NOTE 2 Technology evaluation will always involve data storage for later, offline processing. However, with scenario evaluations, online transactions might be simpler for the tester – the system is operating in its usual manner and storage of samples, although recommended, is not necessary.

online testing
execution of enrolment and matching at the time of image or signal submission

NOTE 1 In online evaluations, the experimenter may decide not to retain biometric samples, reducing storage requirements and in certain cases ensuring fidelity to real-world system operations. However, retention of samples in online tests is recommended for auditing and for subsequent offline analysis.

NOTE 2 Testing a biometric system will involve the collection of input images or signals, which are used for template generation at enrolment and for calculation of matching scores at later attempts. The images/signals collected can be used immediately either for an online enrolment, verification, or identification attempt, or may be stored and used later for offline enrolment, verification, or identification.

performance interoperability matrix
an m by n matrix in which the value contained in each cell, (x , y), gives a performance metric (such as FRR at a fixed FAR or FNMR at a fixed FMR) associated with enrolment using biometric product x and verification using biometric product y.

presentation
submission of a single biometric sample on the part of a user

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

sample
user's biometric measures as output by the data capture subsystem

EXAMPLE Fingerprint image, face image and iris image are samples.

scenario script
a script utilized by an administrator in the direction of a user during enrolment and recognition transactions

similarity score
measure of the similarity between features derived from a sample and a stored template

NOTE 1 A match or non-match decision may be made according to whether this score exceeds a decision threshold.

NOTE 2 As features derived from a presented sample become closer to the stored template, similarity scores will increase.

target population
set of users of the application for which performance is being evaluated

template
model of the user's stored reference measure based on features extracted from enrolment samples

NOTE The reference measure is often a template comprising the biometric features for an ideal sample presented by the user. More generally, the stored reference will be a model representing the potential range of biometric features for that user.

test organization
functional entity under whose auspices the test is conducted

test subject
user whose biometric data is intended to be enrolled or compared as part of the evaluation

transaction
sequence of attempts on the part of a user for the purposes of an enrolment or verification

NOTE There are two types of transactions: enrolment sequence, resulting in an enrolment or a failure to enrol; or a verification sequence resulting in a verification decision.

user
person presenting biometric sample to the system

verification
application in which the user makes a positive claim to an identity, features derived from the submitted sample are compared to the enrolled template for the claimed identity, and an accept or reject decision (and possibly a match similarity score) regarding the identity claim is returned

verification decision
determination of the validity of a user's claim to identity in the system

zero-effort impostor attempt
attempt in which an individual submits his/her own biometric characteristics as if he/she were attempting successful verification against his/her own template, but the comparison is made against the template of another user

Annex A: API Specification

This is the document that was provided to the vendors prior to the test so that they could ensure their products would satisfy all functions required for biometric enrolment and verification associated with ILO SID and so that their products could be easily integrated into the test harness.

API Specification

ILO SID Biometric Interoperability Test

October 2007

Participants in ISBIT shall be required to provide Bion Biometrics, the ILO designated test lab, with an Application Programming Interface (API) that complies with the specifications detailed in this document.

Note: All BDIRs and raw BMP images produced by products for ISBIT will become the property of Bion Biometrics, which will safeguard them in accordance with all relevant privacy legislation under the terms of the personal information release forms signed by test subjects. In order to resolve interoperability issues or to support future offline tests, raw BMP images and/or corresponding SID-0002 BDIRs produced by each vendor's product may be anonymously shared.

A.1 API and Platform Requirements

The API shall be submitted in the form of a compiled Win32 dynamic link library (DLL) which runs on Microsoft Windows XP SP2. The test software that imports the API functions is written in C# on the Microsoft .NET Framework 2.0 using the DllImport attribute.

The API specified by this document shall be implemented in a single base DLL file with the filename '**`isbitapi.dll`**'. Additional dynamic/shared library files may be submitted that support this base library file (i.e. the base DLL may have dependencies implemented in other libraries).

API functions specified to be used during both online and offline testing (**`Enrol`**, **`VerifyProcess`**, and **`VerifyMatch`**) shall not use any interactive mechanisms such as graphical user interface (GUI) calls, or anything requiring terminal interaction including calls to "standard input" or "standard output." These functions shall also run without the presence of the participant's biometric sensor and device drivers.

The API provided must not include multiple "modes" of operation, or algorithm variations.

The API shall access only that system memory that it allocates or that corresponds to the provided inputs and outputs. Furthermore, the API shall not communicate with any external processes, devices, or computers except those required for biometric capture. Modern desktop PCs with USB 2 and Firewire ports will be used for biometric capture in the lab.

A.1.1. Installation

The API should install easily, and shall be executable on any number of machines without requiring additional hardware-based license control procedures. It is recommended that the API be installable using simple file copy methods, and not require the use of a separate installation program.

A.1.2. Documentation

Complete documentation of any functionality or behaviour beyond what is specified in this document should be provided.

A.1.3 Speed

On average, an **`Enrol`** operation should take no more than 7 seconds, a **`VerifyProcess`** operation should take no more than 1 second, and a **`VerifyMatch`** operation should take no more than 10 milliseconds to complete (using a 3 GHz Pentium IV).

A.2 API Function Calls

A.2.1 ImageSize

C Prototype

```
int _stdcall ImageSize();
```

Description

Returns the byte size of each uncompressed BMP image captured. The return value will be used by the calling application (test harness) to allocate an appropriately sized image buffer for the **`Capture`** function of the same product.

Parameters

None.

Return Values

The byte size of each raw fingerprint image in uncompressed BMP format.

A.2.2 ITemplateMaxSize

C Prototype

```
int _stdcall ITemplateMaxSize();
```

Description

Returns the maximum byte size of an intermediate template that could be returned by the **`VerifyProcess`** function of the same product. The return value will be used by the calling application (test harness) to allocate an appropriately sized buffer for the **`VerifyProcess`** function of the same product.

Parameters

None.

Return Values

Buffer size for the **`iTemplate`** parameter of **`VerifyProcess`**.

A.2.3 CaptureInit

C Prototype

```
int _stdcall CaptureInit(const int showGUI);
```

Description

Initializes biometric device before subsequent calls to **`Capture`**. Some devices require a perceptible duration for the automatic initialization and/or calibration of the sensor before running online capture transactions. The test harness will call this function once before each transaction consisting of multiple **`Capture`** attempts for enrolment or verification. This function will not be used during offline testing.

Parameters

`showGui` (input): If the API provides an on-screen indicator via a window or GUI during the initialization period, a value of zero (0) will suppress the indication.

Return Values

0 Success
-1 Failed to Initialize
If initialization for capture is successful, or if no initialization is required by the product, the API should return zero (success) when this function is called.

A.2.4 CaptureEnd

C Prototype

```
int _stdcall CaptureEnd();
```

Description

Provides the opportunity for the API to perform any 'housekeeping chores' or resource de-allocation that may be required at the conclusion of a capture transaction. The test harness will call this function once after each transaction consisting of multiple **`Capture`** attempts for enrolment or verification. This function will not be used during offline testing.

Parameters

None.

Return Values

0 Success
-1 Failed to End Capture
If the capture session is shutdown successfully, or if no shutdown is required, the API should return zero (success) when this function is called.

A.2.5 Capture

C Prototype

```
int _stdcall Capture(
                    const unsigned char finger,
                    const unsigned char purpose,
                    unsigned char       *image);
```

Description

Displays a window or GUI to prompt for placement of the finger corresponding to the finger parameter, and capture a single raw fingerprint image from the biometric device for either enrolment or verification as specified by the **`purpose`** parameter. This function will not be used during offline testing.

This function will be called once for each finger placement. Multiple placements will not be permitted during a single capture call, and a BMP image of the same size (as specified by the return value of the **`ImageSize`** function) must always be output, even if it is blank. If finger placement is automatically detected by the API, it must exit once the finger is removed from the sensor or the image has been acquired. If the API deems the image as unsuitable for the **`purpose`** indicated, it shall output the image and a return value of -1 (Failed to Acquire).

If, after 12 seconds, the administrator determines that the API fails to detect a legitimate finger placement, a button shall be provided in the GUI to allow the administrator to cancel the current capture operation, outputting an image and a return value of -2 (Cancelled by User). If the capture operation is cancelled, the presentation will not count as a failure to acquire by the test control software, and the image will be processed for the **`purpose`** indicated.

Parameters

`finger` (input): A value from 1 to 10 corresponding to a valid finger position from SID-0002 or ANSI/NIST-ITL 1-2000, table 5.

`purpose` (input): A value of zero (0) will indicate a capture for the purpose of enrolment, while a non-zero value will indicate a capture for the purpose of verification.

`image` (output): The raw fingerprint image in uncompressed BMP format. A buffer will be allocated by the calling application to the size returned by the **`ImageSize`** function of the same product.

Return Values

0	Success
-1	Failed to Acquire
-2	Cancelled by User

A.2.6 Enrol

C Prototype

```
int _stdcall Enrol(
                    const unsigned char fingerPrimary,
                    const unsigned char fingerSecondary,
                    const unsigned char *imagePrimary1,
                    const unsigned char *imagePrimary2,
                    const unsigned char *imagePrimary3,
                    const unsigned char *imageSecondary1,
                    const unsigned char *imageSecondary2,
                    const unsigned char *imageSecondary3,
                    unsigned short      *birSize,
                    unsigned char       *bir);
```

Description

This function shall attempt to enrol both primary and secondary fingers as an SID-0002 BDIR using up to three uncompressed BMP images for both the primary and secondary fingers captured on the same biometric device. This function will be used for both online and offline testing.

An SID-0002 conformant BDIR should always be output. Therefore, if either the primary or the secondary finger could not be enrolled from the input images, the enrolled finger shall be designated as the primary fingerprint template and the secondary fingerprint template shall be 'unenrolled'. (see SID-0002 section 5.1.1 and Annex B) If neither the primary nor the secondary set of images could be enrolled, both the primary and secondary fingerprint templates of the BDIR shall be 'unenrolled'. Return values -1, -2, and -3 will indicate that the enrolment of a different finger is required by the test harness.

Parameters

fingerPrimary (input): A value from 1 to 10 corresponding to a valid finger position from SID-0002 or ANSI/NIST-ITL 1-2000, table 5.

fingerSecondary (input): A value from 1 to 10 corresponding to a valid finger position from SID-0002 or ANSI/NIST-ITL 1-2000, table 5.

`imagePrimary1, imagePrimary2, imagePrimary3` (input): Raw uncompressed BMP images from the same product corresponding to the finger identified by **`fingerPrimary`**. May be set to null by the calling application.

`imageSecondary1, imageSecondary2, imageSecondary3` (input): Raw uncompressed BMP images from the same product corresponding to the finger identified by **`fingerSecondary`**. May be set to null by the calling application.

`birSize` (output): The size of the BDIR in bytes.

`bir` (output): SID-0002 BDIR containing two fingerprint minutiae templates. A 566-byte buffer will be allocated by the calling application. The **`birSize`** parameter will specify the actual size.

Return Values

0 Success
-1 Failed to Enrol Primary
-2 Failed to Enrol Secondary
-3 Failed to Enrol Primary and Secondary
-4 Unknown Image Format

A.2.7 VerifyProcess

C Prototype

```
int _stdcall VerifyProcess(
                        const unsigned char   *image,
                        unsigned int        *iTemplateSize,
                        unsigned char         *iTemplate);
```

Description

This function will process an input image (captured from the same product) into an intermediate (or proprietary) template to be used as an input to the same product's **`VerifyMatch`** function. This function will be used for both online and offline testing. This function is provided to enhance matching speed in the offline tests when many matches will be performed. It is assumed that in an operational verification, the system performing the verification would receive a live sample of the seafarer's fingerprint to compare with the BDIR read from the SID. Since the live sample would not have to be converted to an SID-0002 conformant format, this function allows vendors to use a proprietary format for those verification images if they so choose.

Parameters

`image` (input): Raw uncompressed BMP image from the same product.

iTemplateSize (output): The size of the intermediate template in bytes.

`iTemplate` (output): Intermediate template to be used as input to VerifyMatch function. A buffer will be allocated by the calling application to the size returned by the **`ITemplateMaxSize`** function of the same product.

Return Values

0 Success
-1 Failed to Process Image
-2 Unknown Image Format

A.2.8 VerifyMatch

C Prototype

```
int _stdcall VerifyMatch(
                    const unsigned int      iTemplateSize,
                    const unsigned char     *iTemplate,
                    const unsigned short    birSize,
                    const unsigned char     *bir,
                    const int               useSecondary,
                    unsigned short          *score,
                    int                     *match);
```

Description

This will attempt to compare an intermediate template from the same product with either the primary or the secondary template within the input SID-0002 BDIR. If the return value is non-zero, then the match and score parameters will be ignored. This function will be used for both online and offline testing.

Parameters

iTemplateSize (input): The size of the intermediate template in bytes.

iTemplate (input): Intermediate template from the same product.

`birSize` (input): The size of the BDIR in bytes.

`bir` (input): SID-0002 BDIR containing two fingerprint minutiae templates.

`useSecondary` (input): A non-zero value shall indicate that the intermediate template should be matched with the secondary template of the SID-0002 BDIR. A value of zero indicates that the intermediate template should be matched with the primary template of the SID-0002 BDIR.

`score` (output): A similarity score resulting from the comparison of the intermediate template with the primary or secondary template of the SID-0002 BDIR. The range of scores should be from a perfect non-match value of 0 (zero) to a perfect match value of 65,535.

`match` (output): A successful match (as determined by the internal threshold of the product) shall be indicated by a non-zero value, while an unsuccessful match will result in a value of zero (0).

Return Values

0	Success
-1	Failed to Process Intermediate Template
-2	Failed to Process BDIR
-3	Failed to Process Intermediate Template and BDIR

A.2.9 Release

C Prototype

```
void _stdcall Release();
```

Description

Frees all resources allocated by the API through prior function calls.

Parameters

None.

Return Values

None.

Annex B: ISBIT – 4 Methodology Description

This is the document that was provided to the vendors prior to the test so that they would understand the conditions under which their product would have to operate in the test and the criterion for its successful addition to the ILO list of approved products.

Methodology

ILO SID Biometric Interoperability Test

October 2007

B.1 Terms and Definitions

For the purposes of this document, the following terms and definitions apply:

administrator
person performing the testing or enrolment, recording test data and/or monitoring the crew

attempt
submission of one (or a sequence of) biometric samples to the system
NOTE An attempt results in an enrolment template, a matching score (or scores), or possibly a failure–to-acquire.

Biometric Identification Record (BDIR)
refers to a ILO SID-0002 conformant data record containing up to two fingerprint minutiae templates

crew
set of test subjects gathered for an evaluation

detection error trade-off (DET) curve
modified ROC curve which plots error rates on both axes (false positives on the x-axis and false negatives on the y-axis)

enrolment
application in which the user is processed by a system in order to generate and store an enrolment template for that individual

enrolment attempt
the submission of three enrolment presentations of one finger on the part of a user for the purpose of enrolment in a biometric system

enrolment presentation
the submission of a single biometric characteristic (fingerprint) on the part of a user for the purpose of enrolment

enrolment transaction
sequence of up to 10 enrolment attempts (one per finger) on the part of a user resulting in an enrolment or a failure-to-enrol

experimenter
person responsible for defining, designing, and analyzing the test

failure-to-acquire rate (FTA)
proportion of verification transactions for which the system fails to capture or locate an image or signal of sufficient quality

failure-to-enrol rate (FTE)
proportion of the population for whom the system fails to complete the enrolment process
NOTE The observed failure-to-enrol rate is measured on test crew enrolments. The predicted/expected failure-to-enrol rate will apply to the entire target population.

false accept rate (FAR)
proportion of verification transactions with wrongful claims of identity that are incorrectly confirmed

false match rate (FMR)
proportion of zero-effort impostor attempt samples falsely declared to match the compared non-self template
NOTE The measured/observed false match rate is distinct from the predicted/expected false match rate (the former may be used to estimate the latter).

false non-match rate (FNMR)
proportion of genuine attempt samples falsely declared not to match the template of the same characteristic from the same user supplying the sample
NOTE The measured/observed false non-match rate is distinct from the predicted/expected false non-match rate (the former may be used to estimate the latter).

false reject rate (FRR)
proportion of verification transactions with truthful claims of identity that are incorrectly denied

features
digital representation of the information extracted from a sample (by the signal processing subsystem) that will be used to construct or compare against enrolment templates
EXAMPLE Minutiae coordinates and principal component coefficients are features.

genuine attempt
single good-faith attempt by a user to match their own stored template

guidance
direction provided by an administrator to a test subject in the course of data capture for enrolment or verification
NOTE Guidance is separate from feedback provided by a biometric system or device in the course of
data capture, such as audible or visual presentation queues.

habituation
the degree of familiarity a test subject has with a device

NOTE A test subject with substantial familiarity using a biometric device, such as that gained in the course of employment, is referred to as a habituated test subject.

impostor attempt
see *zero-effort impostor attempt*

intermediate template
biometric sample generated or processed to conform to a vendor's own closed unknown format

interoperability
measure expressing the verification performance associated with the use by vendor A of biometric data conforming to a standard interchange format generated by vendor B or vice versa

match attempt
the submission of three match presentations on the part of a user for matching in a biometric system

match presentation
the submission of a single biometric characteristic (fingerprint) on the part of a user for matching

match transaction
sequence of two match attempts (corresponding with two templates in a BDIR) on the part of a user simulated during offline testing resulting in a verification decision
NOTE If a BDIR only contains a single enrolled template, a match transaction will consist of a single match attempt.

offline testing
execution of enrolment and matching separately from data capture
NOTE 1 Collecting a database of samples for offline enrolment and calculation of matching scores allows greater control over which samples and attempts are to be used in any transaction.
NOTE 2 Technology evaluation will always involve data storage for later, offline processing. However, with scenario evaluations, online transactions might be simpler for the tester – the system is operating in its usual manner and storage of samples, although recommended, is not necessary.

online testing
execution of enrolment and matching at the time of image or signal submission
NOTE 1 In online evaluations, the experimenter may decide not to retain biometric samples, reducing storage requirements and in certain cases ensuring fidelity to real-world system operations. However, retention of samples in online tests is recommended for auditing and for subsequent offline analysis.

NOTE 2 Testing a biometric system will involve the collection of input images or signals, which are used for template generation at enrolment and for calculation of matching scores at later attempts. The images/signals collected can be used immediately either for an online enrolment, verification, or identification attempt, or may be stored and used later for offline enrolment, verification, or identification.

presentation
submission of a single biometric sample on the part of a user

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

sample
user's biometric measures as output by the data capture subsystem
EXAMPLE Fingerprint image, face image and iris image are samples.

scenario script
a script utilized by an administrator in the direction of a user during enrolment and recognition
transactions

similarity score
measure of the similarity between features derived from a sample and a stored template
NOTE 1 A match or non-match decision may be made according to whether this score exceeds a decision threshold.
NOTE 2 As features derived from a presented sample become closer to the stored template, similarity scores will increase.

target population
set of users of the application for which performance is being evaluated

template
model of the user's stored reference measure based on features extracted from enrolment samples
NOTE The reference measure is often a template comprising the biometric features for an ideal sample presented by the user. More generally, the stored reference will be a model representing the potential range of biometric features for that user.

test organization
functional entity under whose auspices the test is conducted

test subject
user whose biometric data is intended to be enrolled or compared as part of the evaluation

transaction
sequence of attempts on the part of a user for the purposes of an enrolment or verification
NOTE There are two types of transactions: enrolment sequence, resulting in an enrolment or a failure-to-enrol; or a verification sequence resulting in a verification decision.

user
person presenting biometric sample to the system

verification
application in which the user makes a positive claim to an identity, features derived from the submitted sample are compared to the enrolled template for the claimed identity, and an accept or reject decision (and possibly a match similarity score) regarding the identity claim is returned

verification decision
determination of the validity of a user's claim to identity in the system

zero-effort impostor attempt
attempt in which an individual submits his/her own biometric characteristics as if he/she were attempting successful verification against his/her own template, but the comparison is made against the template of another user

B.2 References

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

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

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

d. ILO Seafarers' Identity Documents Biometric Testing Campaign Report – Part 2

e. 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)

f. ISO/IEC 19795-1 Biometric Performance Testing and Reporting – Part 1: Principles and Framework

g. ISO/IEC 19795-2 Biometric Performance Testing and Reporting – Part 2: Test Methodologies

h. ISO/IEC 19795-4 Biometric Performance Testing and Reporting – Part 4: Interoperability Performance Testing

B.3 Background

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.

For successful implementation of ILO Convention No. 185, Seafarers' Identity Documents (SIDs) issued in each ratifying State must be able to be used for verifying a seafarer's identity in every other State to which that seafarer travels in the course of his or her duties. Since this represents 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 Seafarers' Identity Documents, which is 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 first ILO Seafarers' Identity Document Biometric Interoperability Test (ISBIT-1) 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, ISBIT-2, was commissioned to investigate what had caused the problems in interoperability. During this study, supplementary guidance to the information contained in ILO SID-0002 was developed in order to provide clarity on certain areas in the standard that were suspected to be the source of problems. After the vendors modified their software in the light of the new guidance information, the images collected in the previous test were used in an offline test with the new software. In this case, all of the major interoperability problems were resolved and three products (A, C, and F) were determined to be interoperable at the ILO required performance threshold of 1% FRR at 1% FAR.

ISBIT-2 was completed in early 2005 and by 2006 there were several new products ready to be tested for use with the ILO SID. In early 2006, the three previously approved products were tested along with six new products in the ISBIT-3 test. The methodology for that test was almost identical to the one described in this document, and it resulted in all nine products being placed on an approved product list published by the ILO. This list is available from the ILO web site at http://www.ilo.org/public/english/dialogue/sector/sectors/mariti/products.pdf

The ILO plans to encourage further interoperability tests whenever there are sufficient requests from the vendor community to have products added to the ILO's list of products mentioned above. The present test, ISBIT-4 is designed to allow any additional desktop products, such as the nine already qualified, to be tested.

B.4 Introduction

This test methodology is designed to determine whether products submitted for testing satisfy the biometric-related requirements of ILO Convention No. 185 and ILO SID-0002. To determine whether products meet the ILO's requirements, two primary biometric functions are performed: enrolment and verification.

During enrolment, a test subject will attempt to enrol a primary and a secondary finger. If necessary, the test subject can try up to all ten fingers to get two fingers enrolled. The test subject is considered enrolled if at least one finger is enrolled.

During verification, the test subject will attempt to match their primary or secondary finger with a BDIR previously enrolled. The test subject is considered verified if either finger is matched. A limited number of genuine comparisons are performed by each test subject during online testing, while exhaustive genuine and impostor comparisons are performed during a subsequent offline test.

The products submitted to the lab will be tested for conformance to ensure that they can produce and read fingerprint templates in the form of the Biometric Interchange Records (BDIRs) defined in Annex B of SID-0002. If they are conformant, then they will be integrated with distributed test software and some preliminary interoperability tests will be run in the lab. During this period, any problems will be reported to the vendors and they will have an opportunity to provide updated software and/or hardware if they can do so within the time constraints of this phase of the test. In some cases, this may involve multiple iterations of the vendor providing software, the lab testing it for conformance and preliminary interoperability, and the vendor making modifications based on the feedback from the lab. In order to simplify integration with the test control software and to allow for both online and offline testing to be conducted, a simple API specification that must be satisfied by the software component of each product will be provided to those companies that indicate potential interest in participating in the test.

Those products that can demonstrate conformance and preliminary interoperability will be used in the second phase. During this phase, approximately 180 people will enrol on each system and attempt to verify multiple times on each system against BDIRs generated by the same system and by other systems. These test subjects will each visit the test lab twice, separated by approximately three weeks. After the online portion of the test, the images collected will be used in an extensive offline set of cross-comparisons to allow all possible combinations of enrol on one system and verify on another to be explored for both genuine and impostor distributions. A set of ROC curves will be generated and the generalized false reject rate (GFRR) at a generalized false accept rate (GFAR) of 1% (G^1) will be computed for each product when verifying against enrolled templates from every enrol product.

The mean of the G^1 values for all nine of the previously approved products will be called $G^{MEAN.}$ The maximum of all the G^1 values of the nine previously approved products will be called G^{MAX}. Since all the previous products are approved for both enrolment and verification functions, G^{MEAN} and G^{MAX} will be computed from 81 separate G^1 values. If a new product or set of products is to be added to the approved list then a new calculation of G_2^{MEAN} and G_2^{MAX} will take place using the G^1 values of these new products along with the G^1 values for the previously approved products. The following conditions must then be satisfied for the new product or products to be approved:

3. G_2^{MEAN} is less than or equal to G^{MEAN} OR G_2^{MEAN} is less than 1%
4. G_2^{MAX} is less than or equal to G^{MAX}

Performing third-party, independent testing of biometric products from several vendors for both enrolment and verification will provide a high level of assurance that systems using successfully tested biometric technology will be able to verify seafarers' identities accurately, provided their SIDs were created with another successfully tested biometric technology.

B.5 General Test Conditions

B.5.1 Environment

This test scenario will be executed in a "normal office environment," under indirect fluorescent lighting. The biometric products will be deployed in accordance with recommendations of the product suppliers.

B.5.2 Order Effects

The order in which the biometric products are used could potentially affect performance due to the reasons listed below. Therefore, the order in which products are used will be randomized for each test subject visit.

- Feedback from one biometric product may affect user behaviour (e.g. finger pressure) on another.
- As each product is used, the user may become habituated to presenting their fingerprint and thus may achieve better results with later products.
- On arriving at the test lab, test subjects could be out of breath (if they have hurried to make their appointment) or have cold hands/fingers (when cold outside), recovering to a more normal state after a few minutes.

B.5.3 Test Team

The test team consists of two members: an Experimenter and Administrator. The Experimenter is responsible for the overall management of the test, ensuring consistency in the guidance provided to the test subjects, and reviewing test results on an ongoing basis to ensure integrity. The Administrator guides each test subject through the enrolment and verification visits, ensures that the test system functions properly, and records test subject and visit information.

B.5.4 Test Control Software

The primary functions of the test control software are as follows:

- Integration with biometric products using the API Specification
- Tracking of test subject information including; test subject ID, year of birth, nationality group, job group, and gender
- Online enrolment and verification
- Offline genuine comparisons
- Offline impostor comparisons
- Fingerprint image and template storage, access, and security
- Data analysis and reporting

B.6 Test Crew

In addition to the biometric products, test subjects are needed during the performance and interoperability test phase.

B.6.1 Solicitation

Test crew must be volunteers who are aware of the purpose of the test and are willing to give up their fingerprints and some limited demographic information as part of being tested. Requests for volunteers are distributed through various community groups with notification of the nature of the test, the period over which it will occur, and the means by which the test crew will be rewarded for their participation. Since ISBIT-3 established a test crew, these individuals shall also be contacted explicitly to encourage their participation in ISBIT-4. If a large percentage of the test crew is the same between both tests, then results are more likely to be reproducible.

People that volunteer will have their initial visit scheduled, and will be shown the privacy and data protection statement during that visit. They will be allowed to keep a copy until they return for their second visit, at which point they will sign it for the second time, indicating they have had time to consider it, seek legal counsel if desired, and are completely satisfied with it. If they decline to sign during the second visit, then they will be deleted from the database immediately.

B.6.2 Visits

Each test subject will make two visits to the test lab for the online component of the performance and interoperability test phase. The first visit will require each test subject to enrol on each enrolment-capable biometric product, and verify multiple times on all biometric products (which passed integration and conformance), while the second visit will be a repeat of the first, approximately three weeks later.

At the time of each test subject's first visit, the administrator will enter the following data into the test control software's database: test subject ID, birth year, gender, nationality group, and job group.

The administrator will demonstrate one correct finger placement on each biometric product, and the test subject will be instructed which sensor to use and which finger to present. To represent supervised operating conditions, the administrator will also, whenever a test subject has problems using a biometric product, provide finger placement and quality guidance based on their experience with the products and any available direct feedback from the biometric product (e.g. moistening the finger if it is too dry).

B.6.3 Privacy

It is expected to retain the fingerprint images and templates for approximately 10 years to allow future testing to make use of existing databases. All of this will be outlined in the privacy and data protection statement that each test subject will review and sign.

B.7 Product Solicitation and Integration

Participation in the ILO SID Biometric Interoperability Test is open to all vendors with biometric products compliant with ILO Convention No. 185 and ILO SID-0002. Since ILO Convention No. 185 will be implemented in up to 148 countries, it is important to include as many biometric products in the tests as possible to ensure global access to solution providers.

All vendors interested in participating in the test are provided with this document and a detailed API specification along with any additional requirements for their products in advance of the test. Since the hardware and software provided are evaluated as a single combined biometric product, each biometric vendor is encouraged to select the biometric product that they believe would be most advantageous to them (for a seafaring population) for the purposes of the test.

Desktop biometric product submissions must include an API that complies with the API Specification, and conforms to the relevant parts of ILO Convention No. 185 and SID-0002.

Once successfully integrated into the test control software, each biometric product will be evaluated for stability and its effects on the stability of other biometric products in the test harness. The cooperation of the product vendor with the test lab will be required to resolve any issues related to integration and conformance. Once these issues are resolved successfully, the product may proceed to the final test phase.

B.8 Conformance

Those biometric products, which can be successfully integrated with the test control software, are required to demonstrate conformance to the relevant parts of ILO SID-0002 before they may proceed to the final test phase. A biometric product must therefore meet certain functional and procedural requirements divided into three categories, Enrolment, Verification, and Nominal Interoperability.

B.8.1 Enrolment

Several enrolment trials will be performed to ensure that each biometric product:

- prompts for placement of all ten possible finger positions by name or other visual indicator
- provides visual feedback of the fingerprint image presented to the sensor
- indicates a failure-to-acquire or failure-to-enrol for fingerprints of insufficient quality
- successfully enrols two fingers if two fingers of sufficient quality are available
- successfully enrols one finger (in the event no other finger is available)
- produces BDIRs that conform to the data format specified in ILO SID-0002 Annex B

B.8.2 Verification

Several verification trials will be performed to ensure that each biometric product:

- prompts for placement of all ten possible finger positions by name or other visual feedback (e.g. if the primary enrolled finger in the BDIR is a right index finger, the product must ask the user to present their right index finger)
- provides visual feedback of the fingerprint image presented to the sensor
- indicates a failure-to-acquire for fingerprints of insufficient quality
- correctly interprets both enrolled and "unenrolled" templates from conformant BDIRs
- indicates a match for most genuine comparisons
- indicates a non-match for most impostor comparisons
- indicates a similarity score as defined in the API Specification

B.8.3 Nominal Interoperability

Each biometric product will be tested for basic interoperability by attempting to verify at least one of the primary or secondary fingers against conformant BDIRs enrolled on each of the other biometric products. The product is considered to have passed a single interoperability test for a particular BDIR from another product if either the primary or the secondary finger is verified within three match presentations.

For a given product to pass this interoperability test overall, it has to pass single interoperability tests with at least 50% of the BDIRs from products other than itself, and 50% of those other products have to successfully pass interoperability tests with the given product's BDIR.

For example, with 11 total products, the BDIR from Product 1 containing the right and left index fingers from a single test subject will be used to attempt a successful verification on each of Products 2 through 11. Similarly, the test subject will attempt a successful verification on Product 1 against the BDIRs from each of Products 2 through 11. For Product 1 to be considered interoperable, at least 50% of ten, or five of these ten single interoperability tests must match when verification is being attempted using Product 1 and at least five must match when verification is being attempted on Products 2 through 11 against the BDIR from Product 1. The process would be repeated for a small group of well-habituated test crew and the average number of passed tests should be at least 50% in both cases.

Any product that fails at this stage will not proceed to the performance and interoperability test phase.

B.9 Performance and Interoperability

The objective of the Performance and Interoperability phase of the ILO SID Biometric Interoperability Test is to determine both native (enrol and verify on the same product) and non-native (enrol and verify on different products) false reject and false accept rates for biometric verification of the test crew over a reasonable period.

The performance component seeks to demonstrate that the biometric technologies being offered in the marketplace are able to provide sufficient accuracy to be reliable for the seafaring population.

The interoperability component seeks the largest combined set of products which can achieve an average false accept rate less than or equal to 1% with an average false reject rate also less than or equal to 1%, as required by ILO SID-0002.

Test subjects will be instructed when to place a finger, and (for most types of sensors) when to remove it. The administrator will consider a presentation as being completed as soon as it is determined that either a) the biometric product indicates a successful capture, or b) the biometric product indicates that it failed to acquire an image of acceptable quality or c) the timeout was reached before the biometric product returned any result. If the test subject removes his/her finger before being instructed to do so, the administrator will cancel and repeat the presentation process from the beginning.

B.9.1 Enrolment

Test subjects are enrolled on each biometric product during both visits in accordance with the requirements stated in ILO SID-0002. During enrolment, a test subject will make two enrolment attempts to enrol a primary and a secondary finger, starting with the right and left index fingers respectively. If an index finger is missing or damaged to the extent that a fingerprint can neither be captured nor enrolled by a biometric product, the test subject will make another enrolment attempt using another finger or thumb according to the presentation order defined in ILO SID-0002, Section 5.1.1.

If none of the subject's ten fingers can be enrolled, then that test subject will be recorded as being unable to enrol on that biometric product. That test subject will not be able to participate in native genuine comparisons on that product during subsequent verifications, although the test subject will still participate in impostor comparisons and non-native genuine comparisons on that product.

All of the output images and BDIRs will be stored in a secure database for subsequent online and offline verifications.

B.9.2 Online Verification

After the enrolment session is complete, each test subject will make a limited number of genuine comparisons against a previously enrolled template on each biometric product. To maintain active participation by test subjects, the match/non-match decision for each attempt will be prominently displayed. In this way, online verification also functions as a controlled data collection of images for all offline genuine and impostor comparisons. Note that the manufacturers of the biometric products will have established initial threshold settings to be used for online verification, and these will determine the match/non-match indications provided here as feedback to the users. Subsequent offline tests will probably determine that other threshold settings are optimal for maximizing interoperability, and these will be the ones used in producing the final G^1 results.

The distributed test software determines the unique finger positions enrolled during that visit for all biometric products (usually two for the right and left index), randomizes the order of products used for match attempts, and randomizes the match attempts for each product.

B.9.3 Offline Verification

Offline testing will allow exhaustive native and non-native genuine comparisons to be performed. That is, every match presentation of a test subject's finger will be matched against every BDIR with the same finger enrolled by the same test subject on all biometric products. Normally this would involve three presentations of each finger and the maximum similarity score of all three will be used as the similarity score for that attempt.

Similarly, exhaustive native and non-native impostor comparisons will be performed offline by attempting to match every match attempt with templates of the same finger for all other enrolled test subjects on all biometric products.

Two-finger match transactions, as defined in ILO SID-0002, will be simulated during offline testing by taking the maximum similarity score of each pair of match attempts using the two fingers from each individual enrolment BDIR. If a BDIR contains only a single enrolled finger, then only a single match attempt will be used to compute the transactional similarity score.

B.10 Data Analysis and Reporting

The final report will include a selection of relevant metrics, but the most important for the decision of the ILO as to which products are considered interoperable will be a single interoperability matrix of G^1 values calculated using the two-finger (six-presentation) offline matching transactions described above. For each possible combination of enrolment biometric product and verification biometric product, an ROC curve will be generated and a threshold score value selected to obtain a GFAR of 1%. The value of GFRR (i.e. G^1) will then be computed at that threshold score and will be entered into the two dimensional interoperability matrix for that enrol/verify combination, as shown in this sample matrix.

FAR = 1.0%	Verify on A	Verify on B	Verify on C	Verify on D
Enrol with A	x.x%	x.x%	x.x%	x.x%
Enrol with B	x.x%	x.x%	x.x%	x.x%
Enrol with C	x.x%	x.x%	x.x%	x.x%
Enrol with D	x.x%	x.x%	x.x%	x.x%

All possible combinations of the existing approved products and one or more of the newly tested products will be considered to determine if the interoperability matrix can satisfy the approval requirements described in the Introduction to this document. The largest group of products that can satisfy those requirements will become the new group of approved products. This list of products will be submitted to the ILO for final approval at their next Governing Body meeting after the test report is completed. Once the list is approved, it will be published on the ILO website for all member states to use in making purchasing decisions.

Annex C: Instructions For Test Administrators

The instructions shown below were provided to the test administrators as part of their training before the data acquisition portion of the test began. The goal was to clearly outline the interactions with each test subject at each visit and to ensure that the test administrators followed the same general script in communicating with the test subjects. The instructions also covered procedures for sensor maintenance and for recovering from failures with sensors or software.

Script for a Single Test Subject at a Single Visit

The test subject will arrive and verify who they are and the scheduled time of their visit. The person that greets them should take care of the pre-capture portion of the visit, either using the PC in the reception area, or, if nobody is available at reception by using one of the client PC's in the test room. After the pre-capture portion of the visit has been dealt with, the test subject should proceed directly to a test station unless the test lab is full, in which case they should be invited to wait in the reception area and offered refreshments.

The pre-capture portion of the visit involves the following:

First Visit

Explain that this is a test of fingerprint devices to ensure that they can work together properly. The goal is to allow seafarers who may be enrolled with one device in their home country to be able to verify their fingerprints on a large variety of fingerprint devices as they travel around the world from one port to another. Ask the test subject if they participated in the previous fingerprint test in 2006. If they did, then ask them for their full name and find their personalized form in the ISBIT-4 binder. Look up the Subject ID from the binder and enter it into IsbitDirector to find the record associated with that subject. If they did not participate in the previous test, then create a new record in IsbitDirector for them and enter their demographic data. In either case, the test subject should be shown the screen of IsbitDirector containing their demographic information and asked to verify that all of it is correct. Any mistakes in spelling of names or other data entry errors should be corrected at that time. Next give the test subject a release form and ask them to review it and fill it out. They should sign the first portion of the form during the first visit, and fill out all fields, but the second signature in the payment section is reserved for the second visit. If they wish to take more time to review the form, they can be given a second unsigned copy of it to take home and review. Note that the demographic data stored in IsbitDirector comprises the following data elements:

First Name

Year of Birth

Ethnic Origin (Choose the best fitting origin from a drop down list)

Manual/Chemical Exposure (None, Light or Heavy are selected from the GUI)

Gender (Male or Female)

Note that the Manual/Chemical exposure will be based on the test subject's job and hobbies. If they do regular heavy labour (construction worker, roofer, etc.) or they use chemicals on their hands (including nurses who must wash hands very frequently) then they get a "Heavy" rating. If they have a hobby that affects the hands, such as rock climbing, they could get a "Light" rating since a hobby is not done as frequently. On the other hand if their hobby is very destructive to the hands (which can usually be seen by looking for scars or other damage) then it could justify a "Heavy". This rating needs to be discussed with the test subject as it is selected.

Next the visit should be started in IsbitDirector and any fingers that are missing or unavailable due to physical injury, bandages, major weeping sores or other issues can be selected in the Unavailable Fingers dialogue.

The final part of the pre-capture portion of the first visit is to take the release form and store it in alphabetical order by last name in the ISBIT-4 binder. This should be done after the test subject has had a chance to review it and fill it out, which they may do while they wait in the reception area.

Second Visit

Ask the test subject for their name and look up their release form in the ISBIT-4 binder. If no release form can be found, then a new one must be filled in (following the process for a first visit except that the form must be completely signed and immediately placed in the binder). Use the Subject ID from the release form or the subject's first name and year of birth to look up their record in IsbitDirector. If IsbitDirector does not show their last visit as ISBIT-4 Visit 1 or if the previous visit was not at least 10 days in the past, then there is a problem and this subject should be referred to the supervising experimenter to determine what has happened. If their record is available then have them review and verify their demographic details. After they have verified their details, start their second visit in IsbitDirector and then select any missing or unavailable fingers in the dialogue box. Note that the missing or unavailable fingers may be different from the previous visit, as minor physical injuries, bandages, etc. can change from one visit to another.

If they are to be taken immediately into the test lab, then they can sign the second portion of the release form and be paid. Alternatively, the payment and signature can be left until the completion of the second visit.

Data Acquisition

The data acquisition process is the same for both the first and second visit. Since some test subjects require more supervision and verbal correction when failing to correctly follow the prompts of the IsbitDirector and IsbitCapture software, it is up to the test administrator to determine how much guidance is required in each case. Similarly, the second visit will usually require less instruction than the first visit since the test subject will be more familiar with the biometric products and with the test procedure so it will require some judgement from the test administrator on when to offer advice. The following guidelines are useful to explain the process that will be followed and what types of feedback and interaction are permitted between the test administrator and the test subject.

Prior to the beginning of each visit, explain to the test subject the general structure of a visit. The following language captures the essence of this explanation.

You will place your fingers multiple times on each of the 12 products being tested. The order in which the products will be used is random, but the test software will explain which product comes next in each case. First you will be asked to place a primary finger and then a secondary finger, usually the left and right index fingers, on each sensor when you enrol on that sensor. Enrolment is the process of capturing an image of your fingerprint and converting it to the standardized minutiae template that will be used with the ILO Seafarers' Identity Document. This minutiae template contains only a map of the critical points from your fingerprint and not an image of the fingerprint. For the purposes of this test, we are also recording the images of the fingerprints, but all of the matching is done using only the minutiae based templates. If either the left or right index finger can't be enrolled after three placements on the sensor, then other fingers will be tried in a predetermined order until a template can be generated. After completing enrolment on all the sensors, the software will prompt you to start verification attempts on each sensor, once again in a random order. In the verification case, each sensor will require six placements of each finger that you enrolled on any of the sensors and you will be told after each one whether it was a match or not. Note that sometimes the system will attempt to match your finger against your own template and sometimes it will be against another person's template, so it will not always result in a match. Your job is simply to give the best fingerprint each time so that if it is a comparison to your own template then a match will be possible.

The software will prompt you and I will tell you when to place the finger on each sensor and when to remove it. Some sensors require calibration prior to finger placement and some sensors do not. If the sensor does perform a calibration, it is important that you do not have your finger on it during calibration, so please do not place your finger until told. Generally, you will be able to place your finger once the sensor capture window appears in the center of the screen and remove it once the image of your fingerprint appears in *the capture summary window in the top left of the screen. Do you have any questions or are you ready to begin?*

At this point, any final questions should be answered and the data acquisition should proceed in the order determined by IsbitDirector. If the test subject does not follow instructions, such as by placing the wrong finger on the sensor, or by placing the finger at an odd angle instead of flat, or by placing the finger during sensor calibration or removing it prior to the image being fully captured, then the current process should be repeated using the "Re-Initialize" or "Re-Acquire" buttons. The test subject should be told what they did wrong and given a chance to correct it. Those sensors that require manual intervention to capture the fingerprint should be triggered as soon as the finger is properly placed and an image is visible. Those sensors that capture automatically should be allowed up to 12 seconds (until the timer bar in IsbitCapture turns red) and then the automatic cancel should be triggered. Such a cancellation is not a reason to re-acquire the current finger presentation unless the timeout was caused by an error on the part of the test subject or the test administrator.

As each product is used, the test administrator should place the current product back in the line of products and pull the new one forward to be placed directly in front of the test subject. A single example of finger placement should be made on the first sensor that is used in order to show how a finger can be placed flat and centered in the middle of the sensor.

If the test subject is consistently pressing too hard or too soft or placing the finger too far in one direction from the center of the sensor, then a verbal suggestion for improving the quality of the fingerprint acquisitions should be made.

If the fingers of the test subject appear to be consistently too moist, then encourage them to wipe them on their clothing to dry them. In extreme cases, they can be offered a dry cloth to help dry their fingers before each presentation. If the fingers appear to be consistently too dry then encourage them to wipe their fingers along their forehead or the side of their nose to acquire some skin oil to moisten their finger. If they are still consistently too dry then allow them to use moisturizing lotion. After applying the lotion allow a minute for it to dry and then offer a dry cloth to remove any excess. Ideally, a decision on whether or not to use moisturizer for a given visit should be made during the enrolment on the first one or two sensors and if moisturizer is to be used then the record for that subject on that visit should be reset and the visit should be restarted after application of the moisturizer so that all products benefit equally from it. If moisturizing lotion is used, then this should be entered into the notes field for the current visit of the test subject.

Any other abnormal occurrences should also be entered in the notes field with as much detail as possible.

Sensor Maintenance

As outlined in the procedures for the beginning and end of each day, all of the 24 sensors used in the test should be wiped with an alcohol swab at the end of each day in a side to side and an up and down motion. During the course of each day, if a sensor becomes extremely dirty (observed either by looking at the surface of the sensor or by noticing significant artifacts on the images displayed during finger capture or by an error occurring due to a dirty sensor during initialization), then it should be cleaned with a dry cloth. If the dirt represents a hygiene issue (such as being residue from the fingers of a test subject with eczema) or if a test subject expresses a hygiene concern due to residue on the sensor, then the sensors should be immediately cleaned again with alcohol and this should be noted in the observation field for the current visit of the current test subject. The sensors should also be cleaned with alcohol during the most convenient break some time between noon and 2:00 PM each day. Allow at least two minutes for the alcohol to completely evaporate before resuming with data acquisition.

Handling of Computer or Biometric Product Errors

The following issues may occur and specific procedures have been developed to handle them.

1. Sometimes, especially on reboot of the PC, a sensor will show up as a new hardware device on Windows and the New Hardware Wizard pops up. The administrator should then reinstall the same driver using the following steps.
   - Tell Windows "Not this time" when it asks if it can search for a driver on the internet. Then click "Next"
   - Select "Specific Location" on the screen when Windows asks where to find the driver and then click "Next"
   - Select "Browse" and move to the location C:\Isbit4\Product where product represents the name of the product for which the driver is being reinstalled. In some cases, a driver sub-folder under the main product folder may need to be selected. For two specific products, the driver must be selected from C:\ProgramFiles\Product.
   - Click Next to install the driver
   - If there is a question about the driver being unsigned select "Continue Anyway"
   - Click "Finish"

2. Sometimes the placement of a finger on a sensor during initialization will cause its capture window to halt with an error message. The error message displayed is different for different sensors, but in all cases, a simple re-initialization is the first solution. If that fails then IsbitCapture should be closed and the capture process for that product restarted from IsbitDirector.

3. If the thread managing capture from a particular product crashes for unspecified reasons, then it may cause IsbitCapture to display an error message. If this is

caught elegantly and gives a specific location and line of code where the error occurred, this should be recorded in the paper logs for future reference. Then the error box should be closed and IsbitCapture should be closed and then restarted. Note that if all associated threads have not terminated properly then the "Capture" button in IsbitDirector will be replaced with an "Abort" button which must be activated prior to restarting the capture process. If the error causes the entire IsbitCapture application to crash and results in an UnhandledException in IsbitDirector, then record any information provided, close the error message, close IsbitCapture (if it is still open) and close IsbitDirector. Then reboot the computer and restart IsbitDirector, choose the current test subject from the list of active visits and restart IsbitCapture. When this is done, it is important to verify that IsbitCapture restarts at the same presentation that it was at when the crash occurred and this presentation should be recaptured using the "Retake" button.

Annex D: Algorithm Modifications

The vendors of the previously approved products D and F indicated that they were interested in testing newer versions of their algorithms. The vendor of Product G, which had produced the two non-conformant templates observed during the data acquisition phase of the test, also agreed to provide a modified algorithm that was expected to fix the sporadic conformance issue. There were therefore three new template generation and matching algorithms to be tested, all of which were already approved (or conditionally approved) products with sensors and data acquisition software that had been tested as part of ISBIT-4. It was therefore possible to test these three modified algorithms simply by substituting them for the template generation and matching components of the three algorithms they were replacing and then regenerating all templates and re-computing all match scores. This had the effect of reproducing the results of the ISBIT-4 test as it would have been if these three modified algorithms had been used during the original test.

When this was done and the results were recomputed, the interoperability matrix generated was as shown below:

Table 10. Modified Algorithm Interoperability Matrix, GFRR at GFAR = 1%

Enrol Product \ Verify Product	A	B	C	D	E	F	G	H	I	J	K	L	MEAN	RANK
A	**1.6**	0.0	0.7	0.9	0.5	0.0	1.7	0.5	0.1	0.5	1.7	0.5	0.7	7
B	1.6	**0.0**	0.7	0.4	0.1	0.1	0.8	0.3	0.1	0.5	1.2	0.5	0.5	3
C	0.9	0.5	**0.7**	0.7	0.7	0.5	0.5	0.5	0.5	0.5	0.7	0.7	0.6	5
D	2.5	1.1	0.9	**0.5**	0.4	0.1	0.9	0.3	0.8	0.8	3.7	0.7	1.1	9
E	1.1	0.1	0.8	0.4	**0.0**	0.1	0.5	0.0	0.0	0.7	0.5	0.3	0.4	1
F	1.1	0.4	0.8	0.8	0.5	**0.0**	0.8	0.3	0.4	0.7	0.7	0.4	0.6	4
G	2.4	1.3	1.6	1.6	1.2	0.7	**1.3**	1.1	1.1	1.6	2.2	1.1	1.4	12
H	2.1	0.7	1.1	0.4	0.4	0.0	0.9	**0.0**	0.3	0.9	1.3	0.4	0.7	6
I	1.2	0.1	0.8	0.4	0.5	0.0	0.8	0.0	**0.0**	0.7	0.7	0.3	0.5	2
J	1.5	0.5	0.7	0.5	0.7	0.5	1.1	0.5	0.5	**0.5**	1.2	0.7	0.7	8
K	2.8	1.1	0.9	1.9	0.9	0.1	1.9	1.2	0.4	1.1	**0.9**	1.3	1.2	11
L	2.2	0.7	1.3	1.2	0.9	0.5	1.5	0.8	1.1	1.1	1.9	**0.8**	1.2	10
MEAN	1.7	0.5	0.9	0.8	0.6	0.2	1.1	0.5	0.4	0.8	1.4	0.6	**0.80**	
RANK	12	4	9	8	5	1	10	3	2	7	11	6		

During this test, the modified version of Product G did not produce any non-conformant templates, indicating that the problem of occasionally creating templates in which two minutiae occupied the same x- and y-position had been eliminated. The overall interoperable performance was indicated by G_2^{MEAN} = 0.80% and G_2^{MAX} = 3.7%. This value of G_2^{MAX} was significantly higher than the value of G^{MAX} = 2.8% observed when the nine previously approved products were tested and therefore the modified version of

Product D did not meet the ILO criteria for inclusion in the approved products list, since it was the combination of enrolment with Product D and verification with the previously approved Product K that generated this maximum value.

The modified version of Product F, however, continued to be one of the best performing products, with a slight (albeit within the uncertainty of the experiment) improvement in its overall performance. The modified version of Product G (one of the 3 new products) also had a slight improvement in interoperable performance, but the main result for it was that it no longer produced non-conformant templates.

The conclusion is that the conditional approval of Product G should become a full and formal approval and that the modified algorithm for Product F should replace the previous version in the ILO list of approved products. The modified version of Product D, however, is not well interoperable with Product K and product D should continue to use its old algorithm for ILO SID purposes. Therefore the new recommended list of approved products will have three new products and only one modified product.

Annex E: Two finger transaction Based DET Curves

Figure 7. Enrol A Verify A – GFRR Versus GFAR

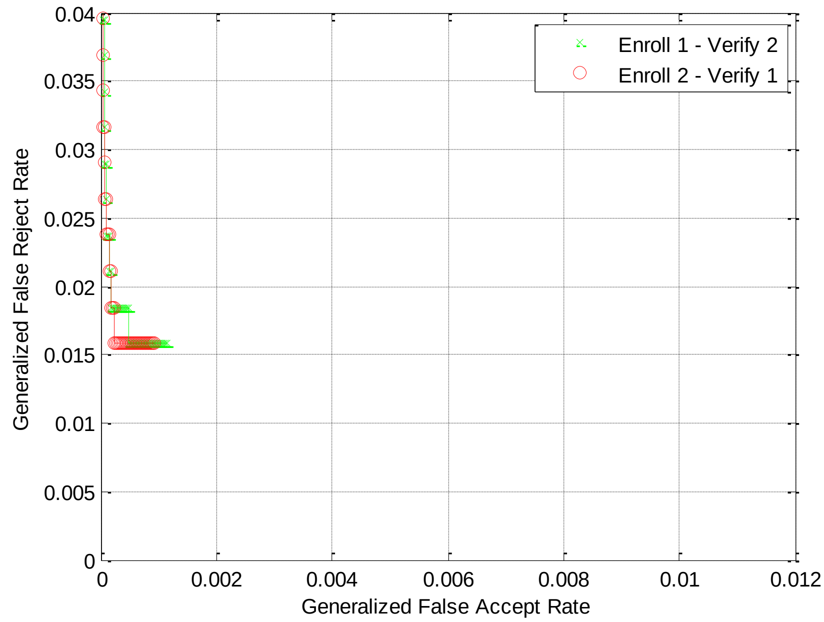


Figure 8. Enrol A Verify B – GFRR Versus GFAR

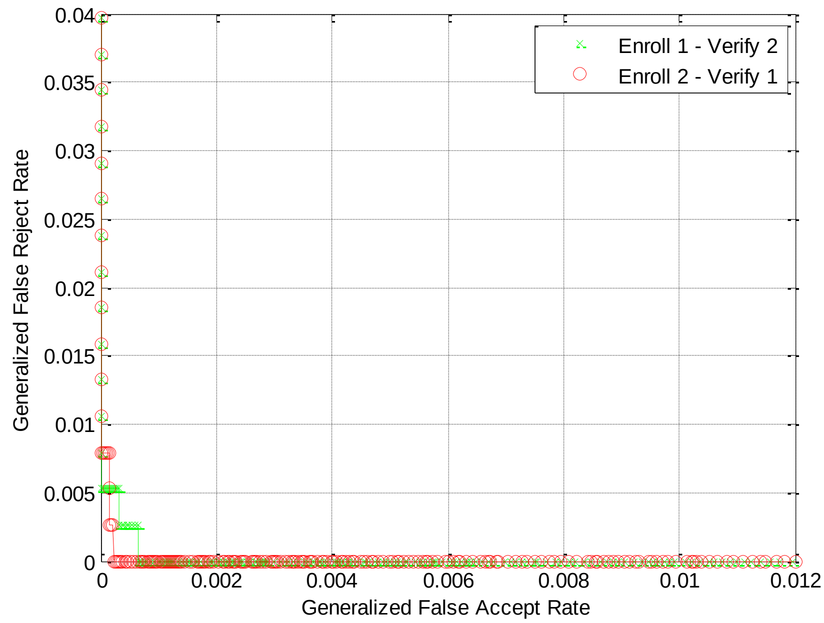


Figure 9. Enrol A Verify C – GFRR Versus GFAR

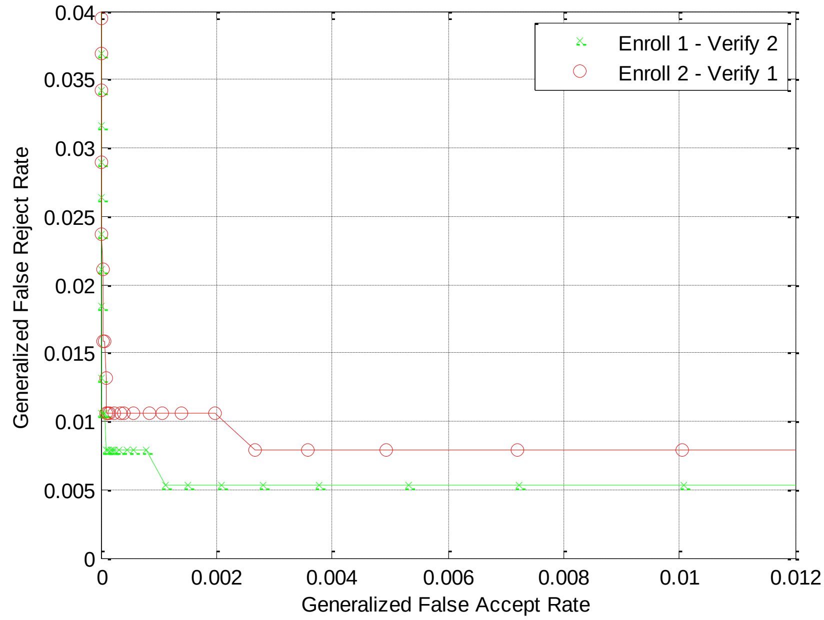


Figure 10. Enrol A Verify D – GFRR Versus GFAR

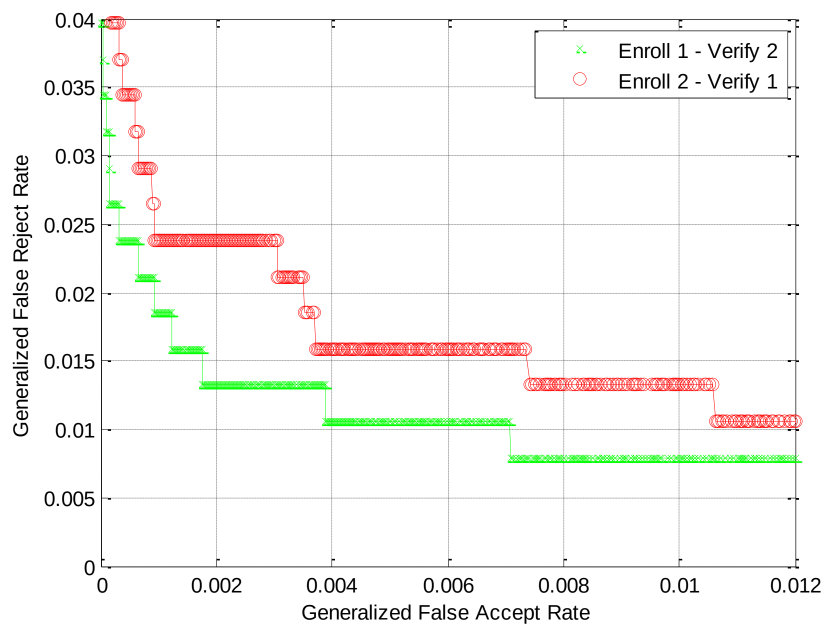


Figure 11. Enrol A Verify E – GFRR Versus GFAR

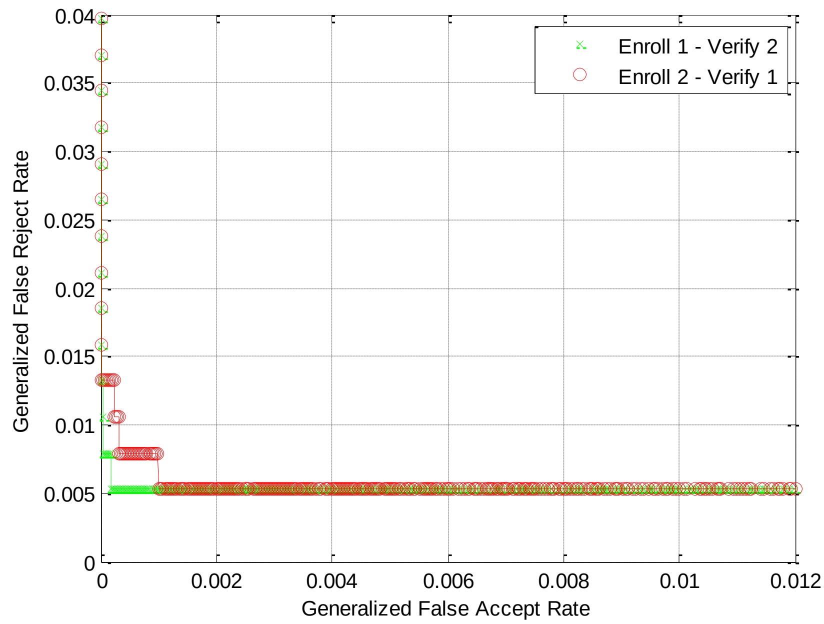


Figure 12. Enrol A Verify F – GFRR Versus GFAR

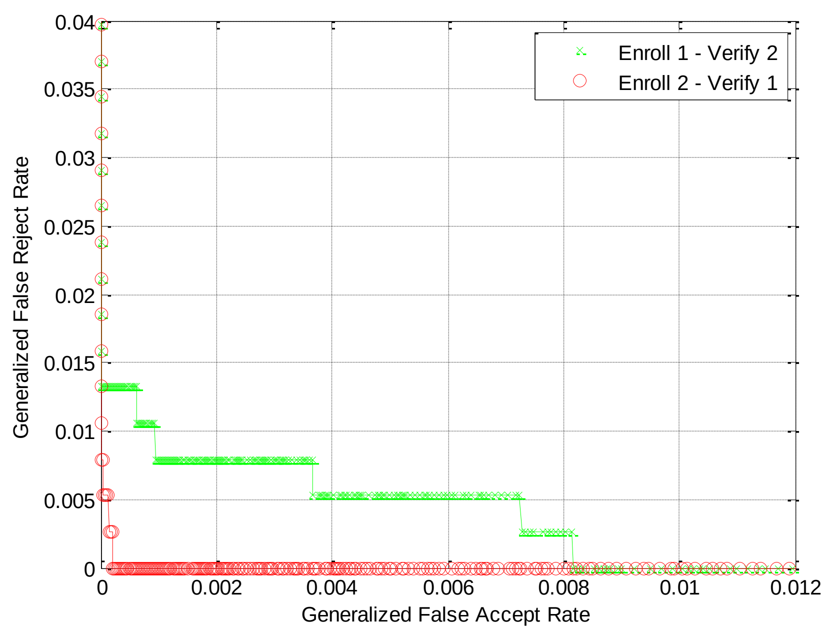


Figure 13. Enrol A Verify G – GFRR Versus GFAR

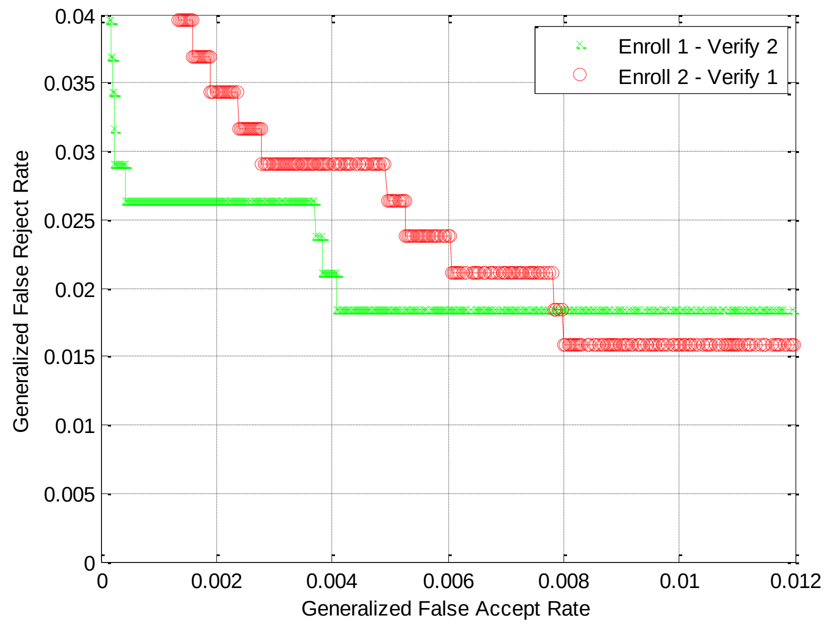


Figure 14. Enrol A Verify H – GFRR Versus GFAR

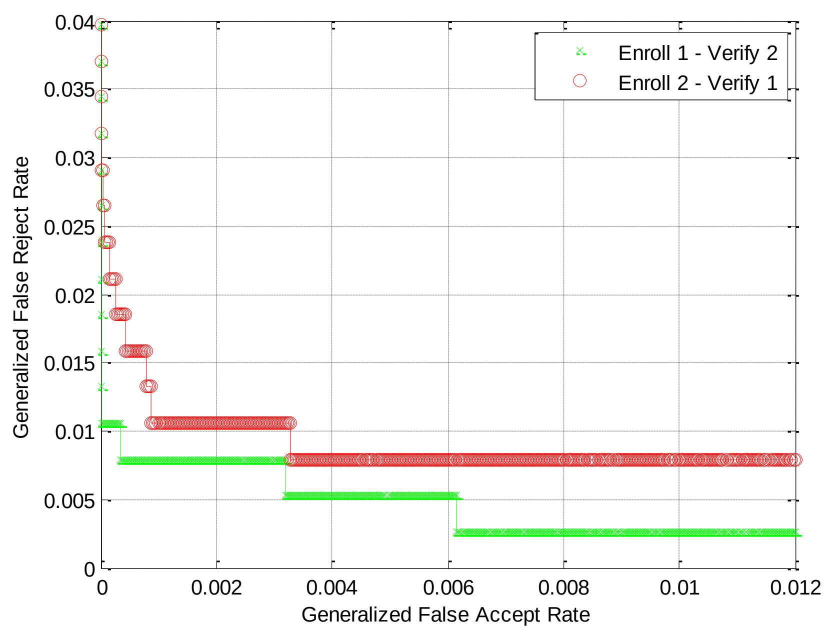


Figure 15. Enrol A Verify I – GFRR Versus GFAR

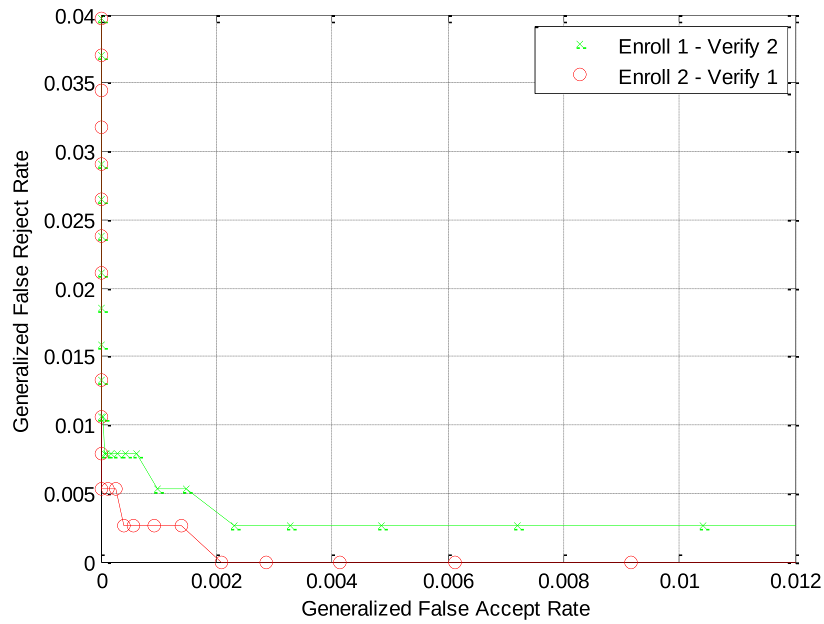


Figure 16. Enrol A Verify J – GFRR Versus GFAR

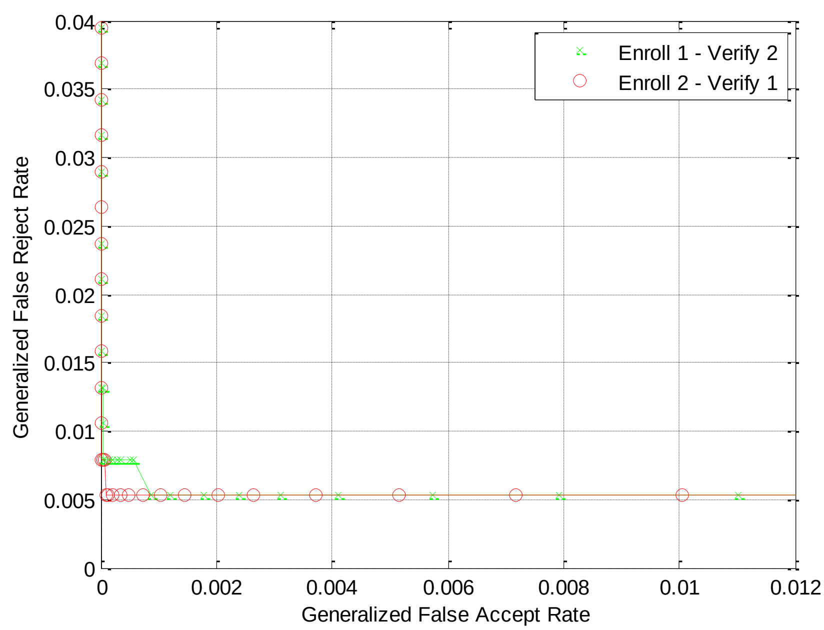


Figure 17. Enrol A Verify K – GFRR Versus GFAR

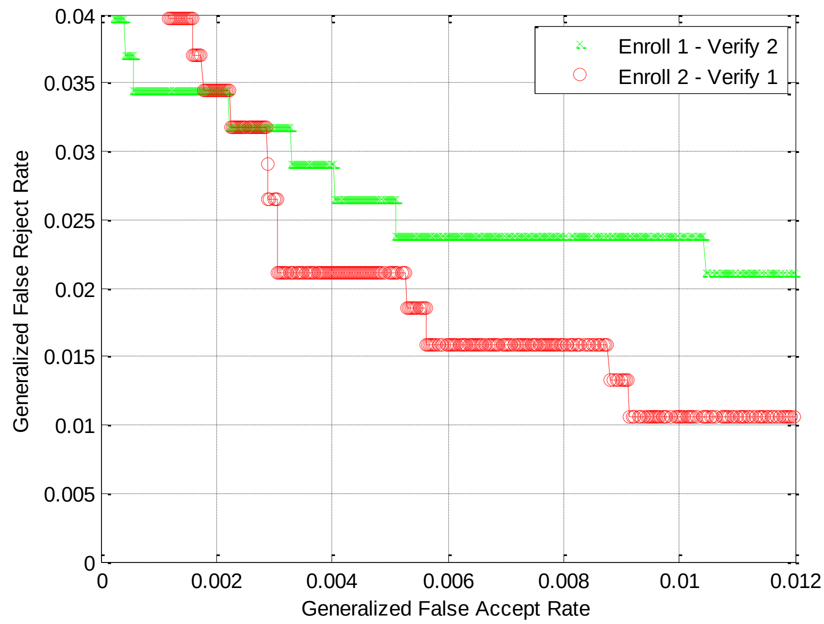


Figure 18. Enrol A Verify L – GFRR Versus GFAR

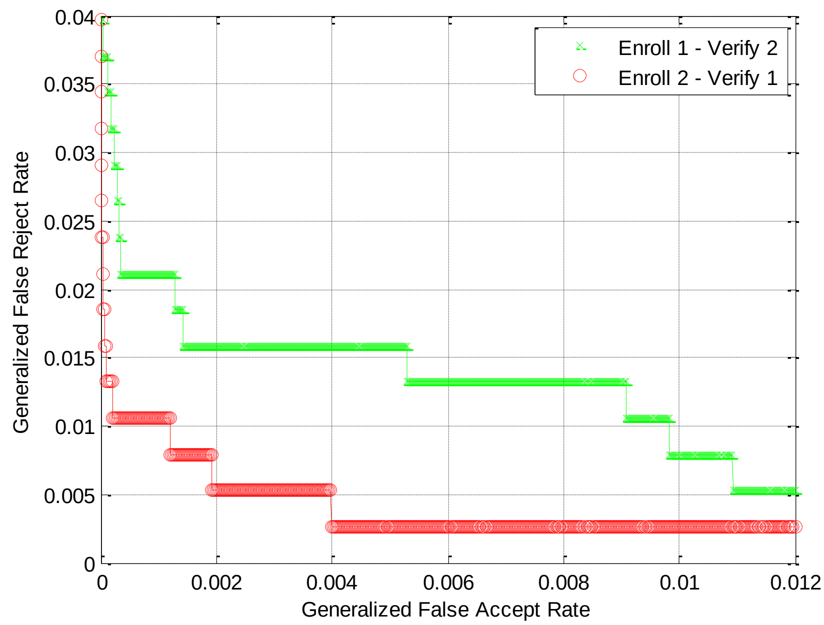


Figure 19. Enrol B Verify A – GFRR Versus GFAR

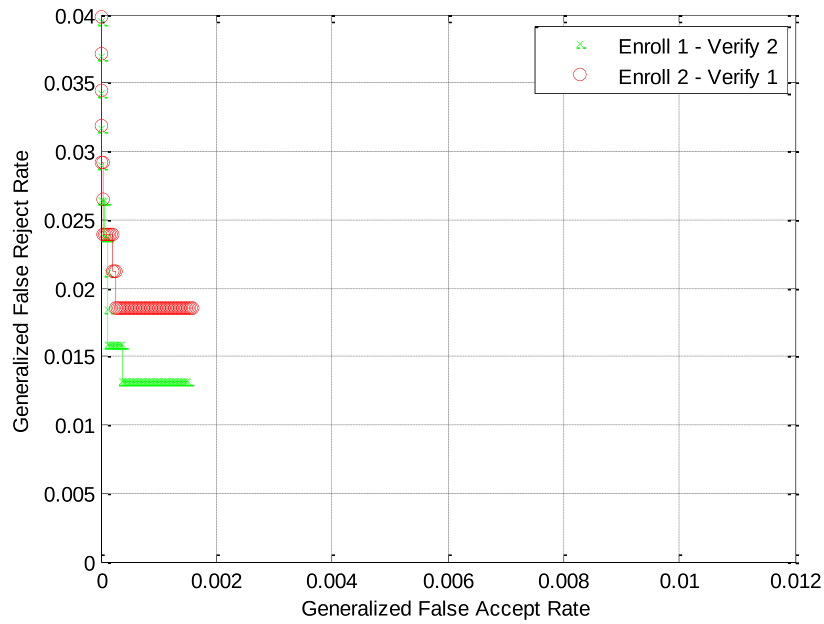


Figure 20. Enrol B Verify B – GFRR Versus GFAR

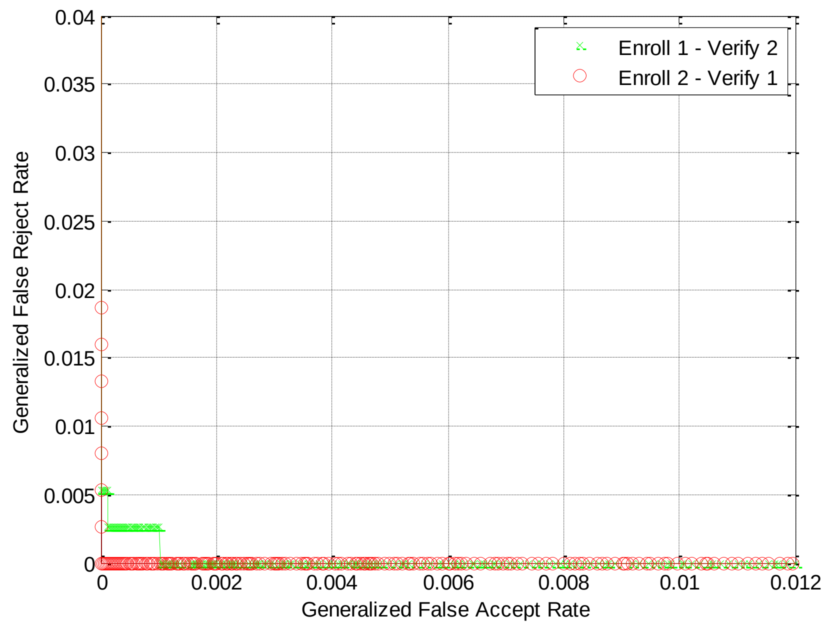


Figure 21. Enrol B Verify C – GFRR Versus GFAR

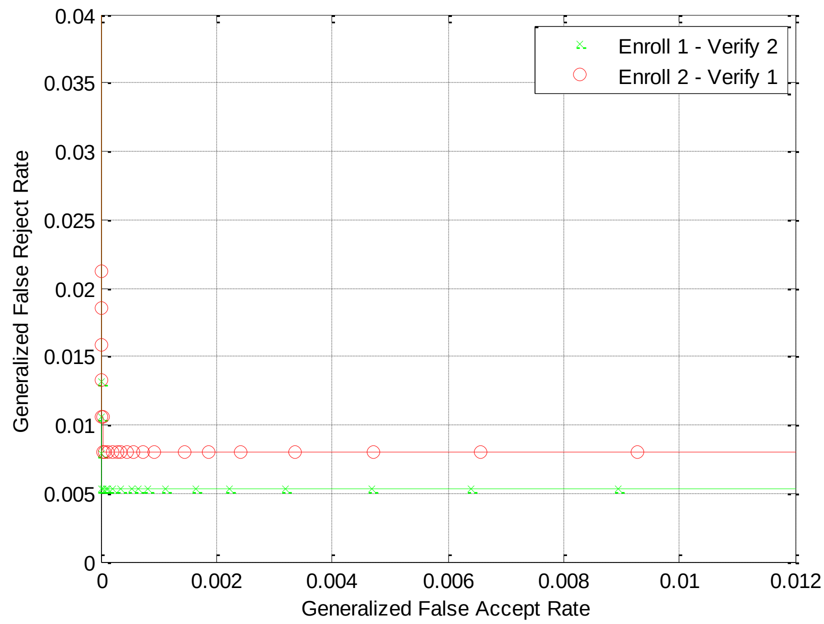


Figure 22. Enrol B Verify D – GFRR Versus GFAR

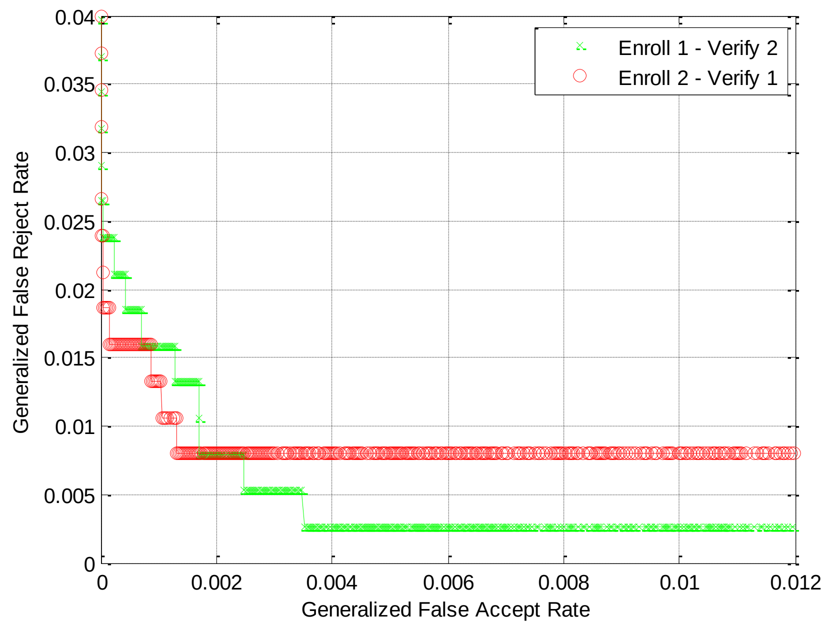


Figure 23. Enrol B Verify E – GFRR Versus GFAR

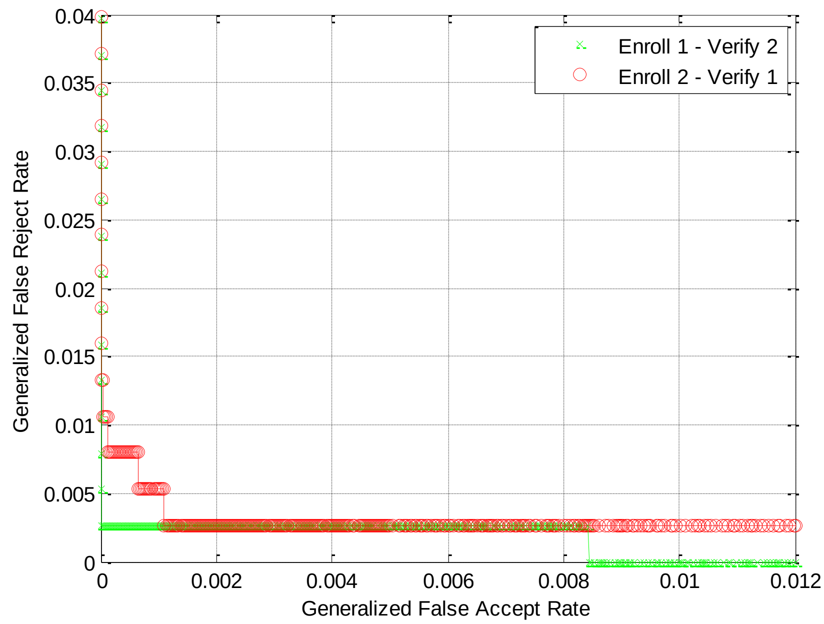


Figure 24. Enrol B Verify F – GFRR Versus GFAR

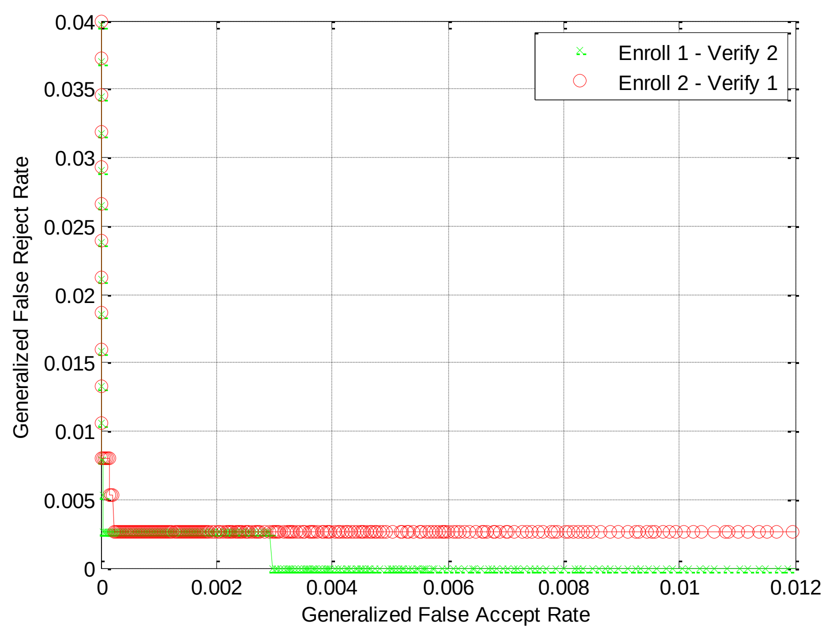


Figure 25. Enrol B Verify G – GFRR Versus GFAR

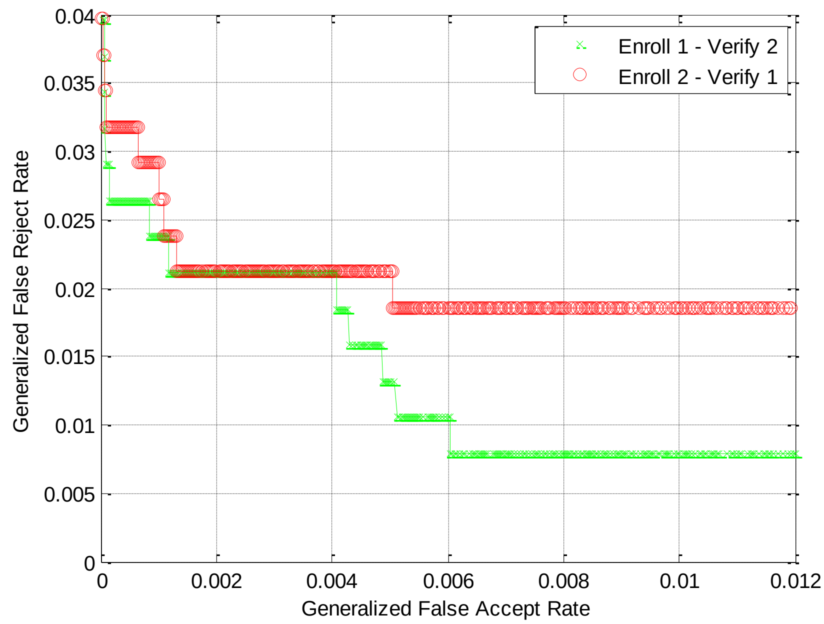


Figure 26. Enrol B Verify H – GFRR Versus GFAR

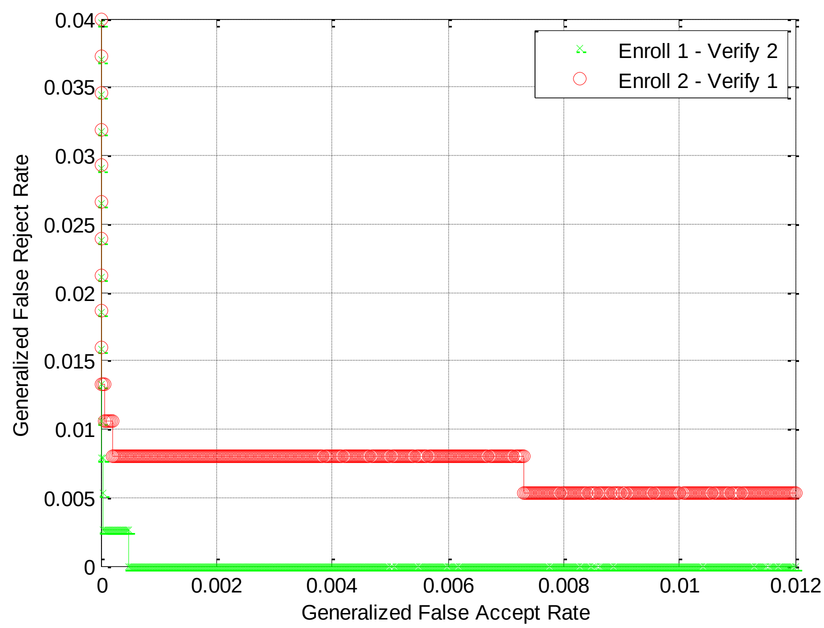


Figure 27. Enrol B Verify I – GFRR Versus GFAR

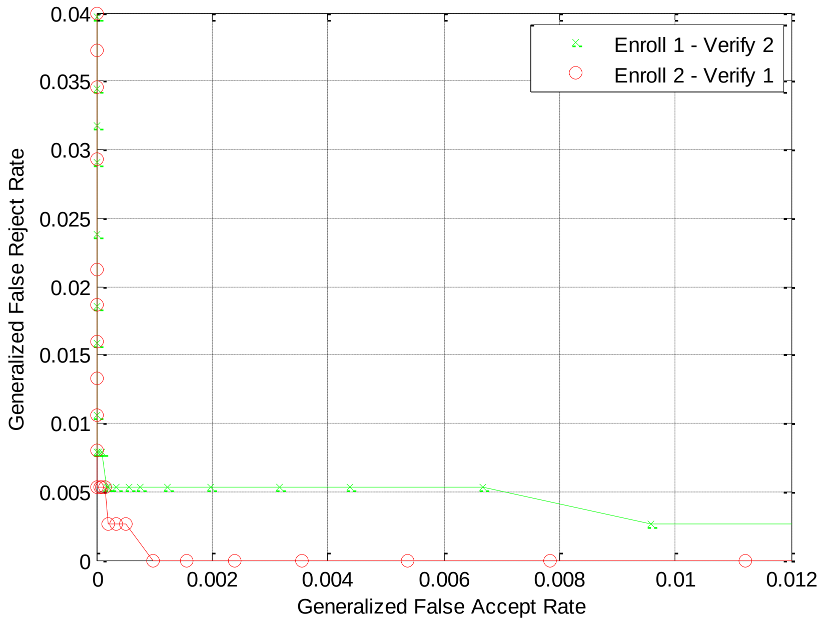


Figure 28. Enrol B Verify J – GFRR Versus GFAR

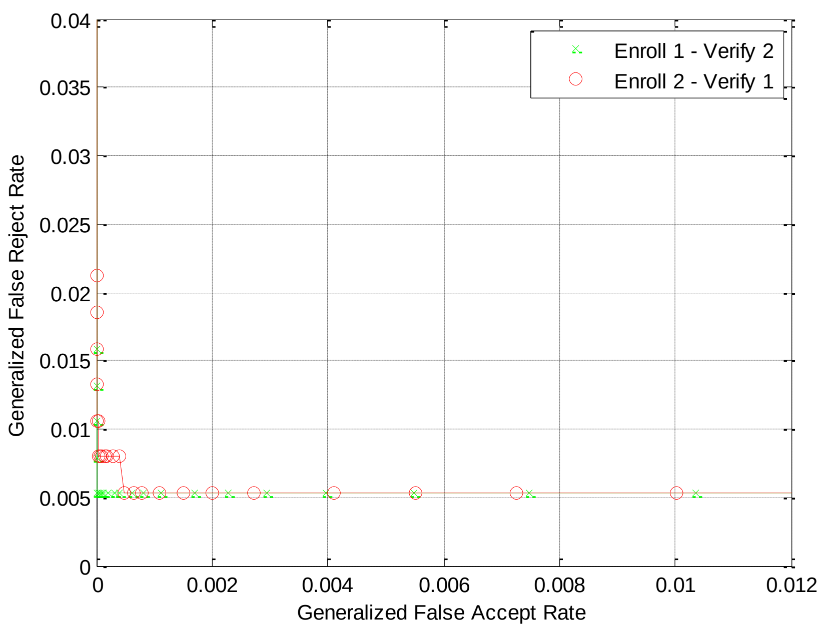


Figure 29. Enrol B Verify K – GFRR Versus GFAR

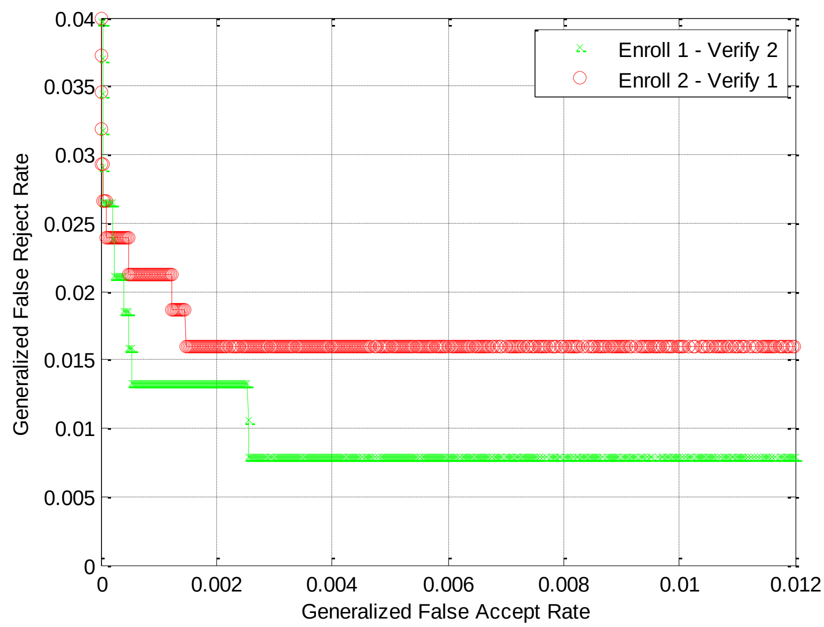


Figure 30. Enrol B Verify L – GFRR Versus GFAR

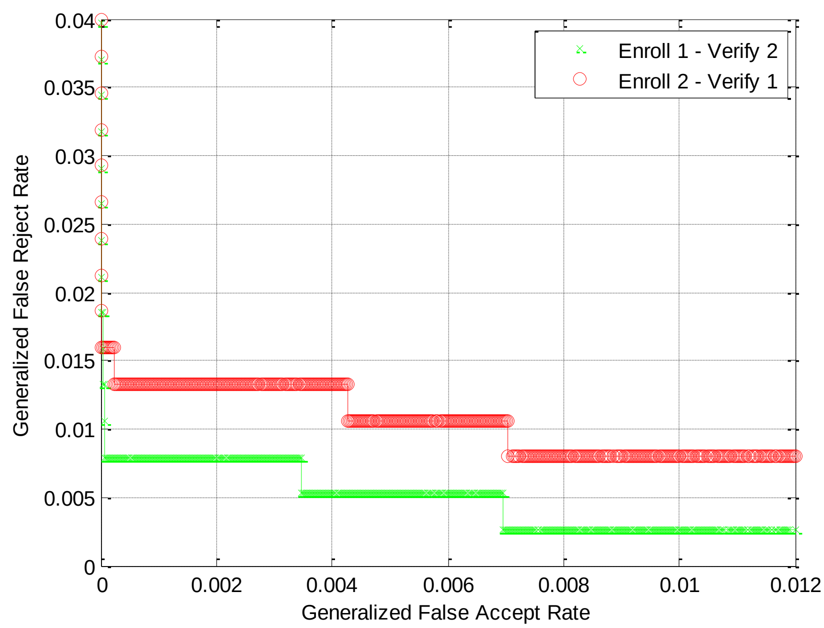


Figure 31. Enrol C Verify A – GFRR Versus GFAR

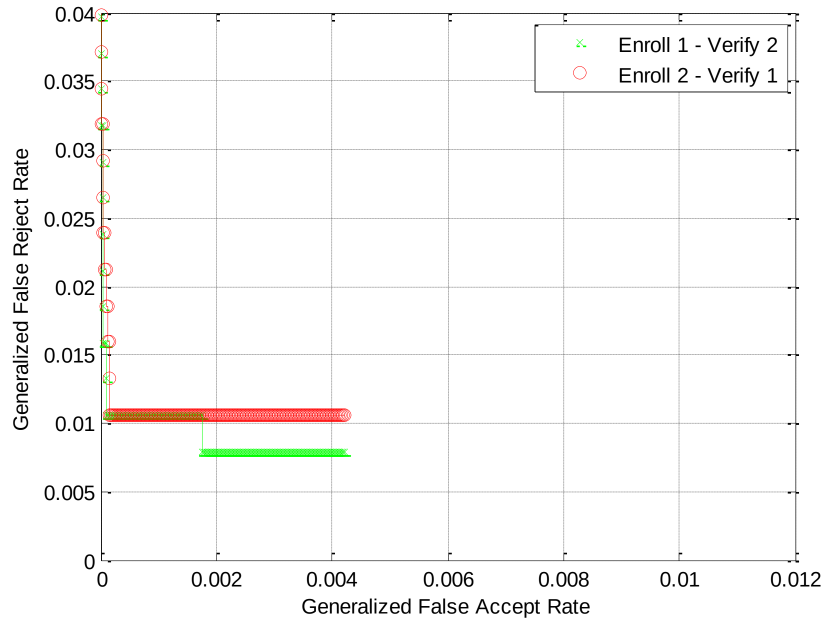


Figure 32. Enrol C Verify B – GFRR Versus GFAR

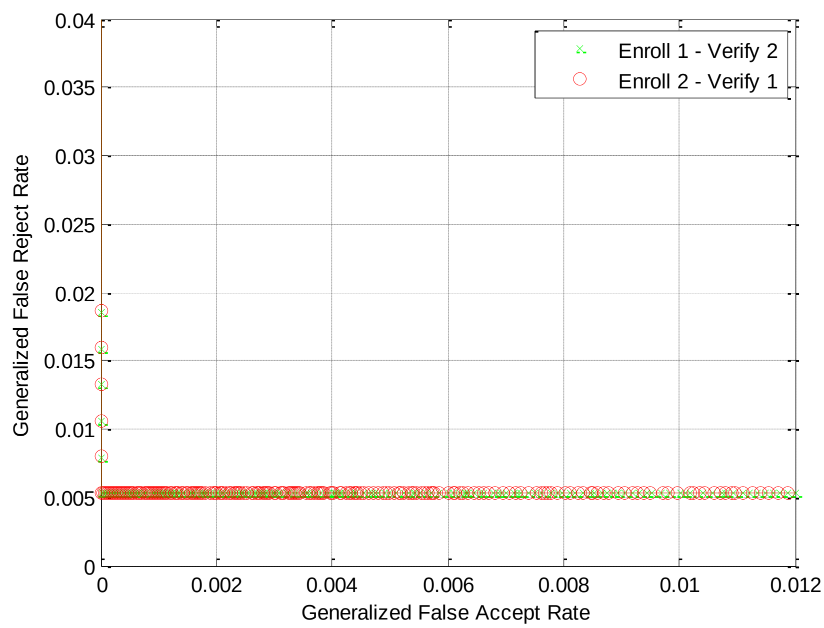


Figure 33. Enrol C Verify C – GFRR Versus GFAR

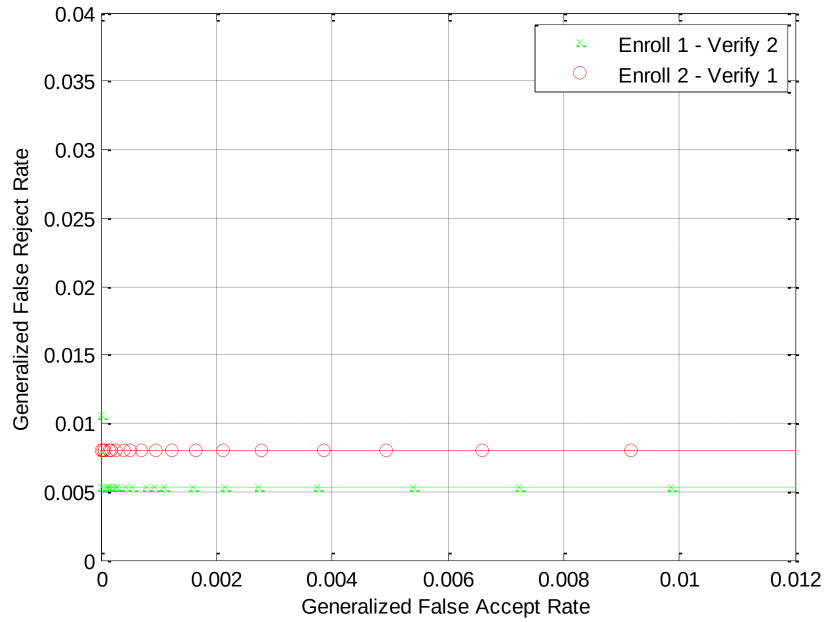


Figure 34. Enrol C Verify D – GFRR Versus GFAR

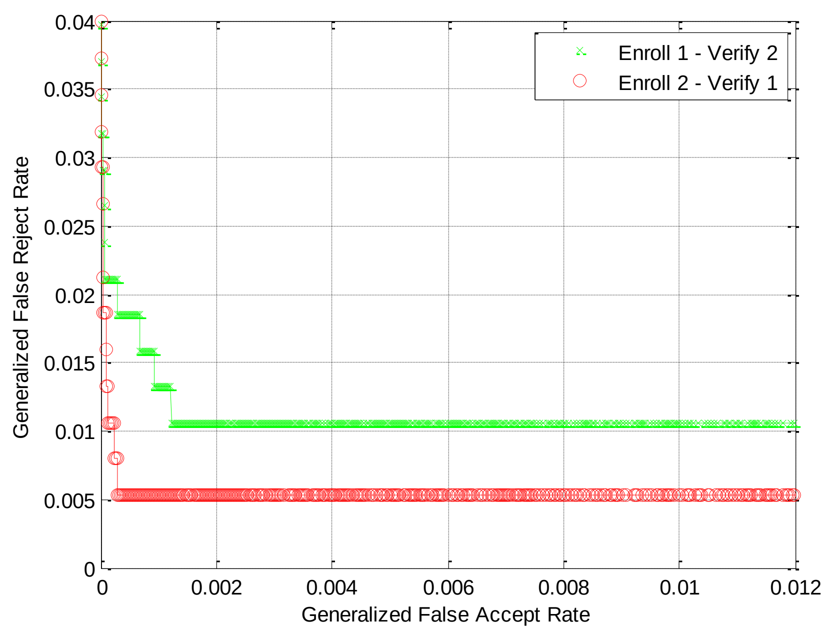


Figure 35. Enrol C Verify E – GFRR Versus GFAR

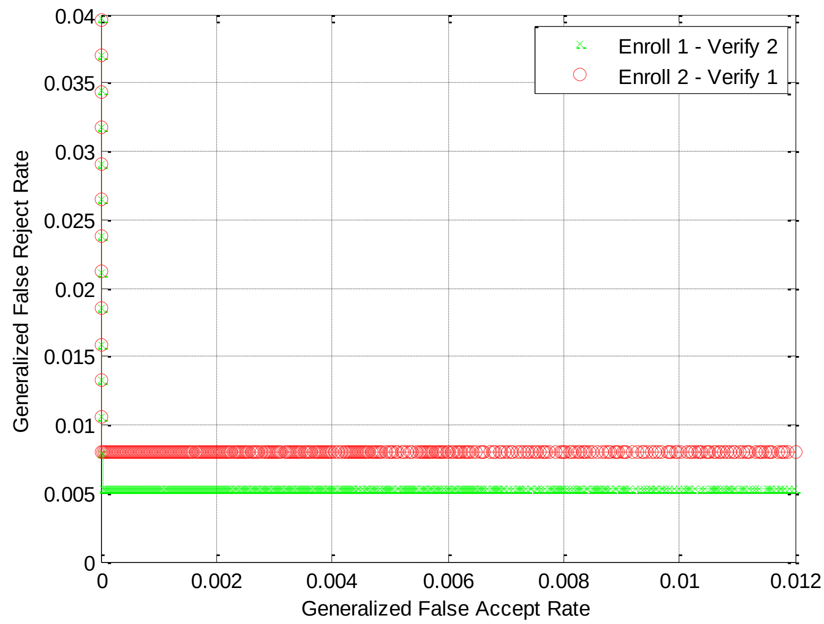


Figure 36. Enrol C Verify F – GFRR Versus GFAR

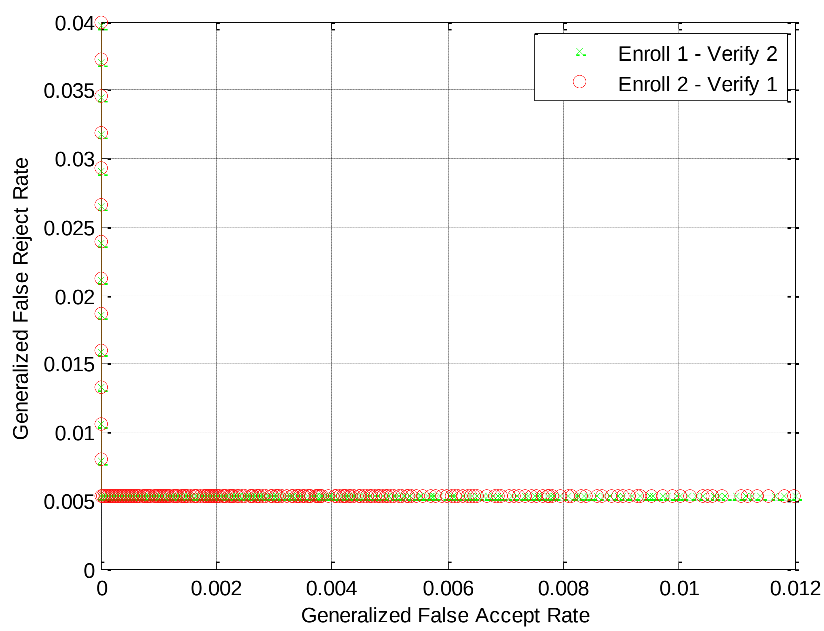


Figure 37. Enrol C Verify G – GFRR Versus GFAR

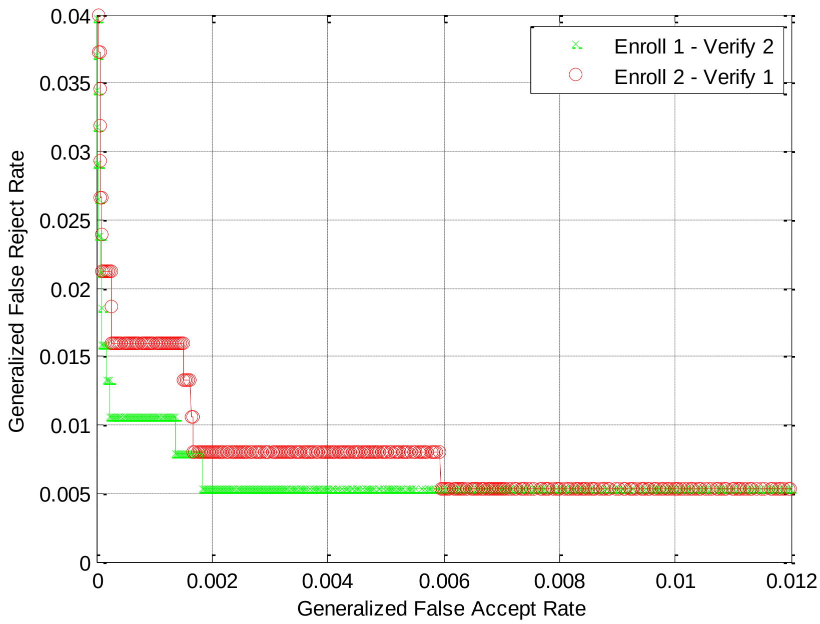


Figure 38. Enrol C Verify H – GFRR Versus GFAR

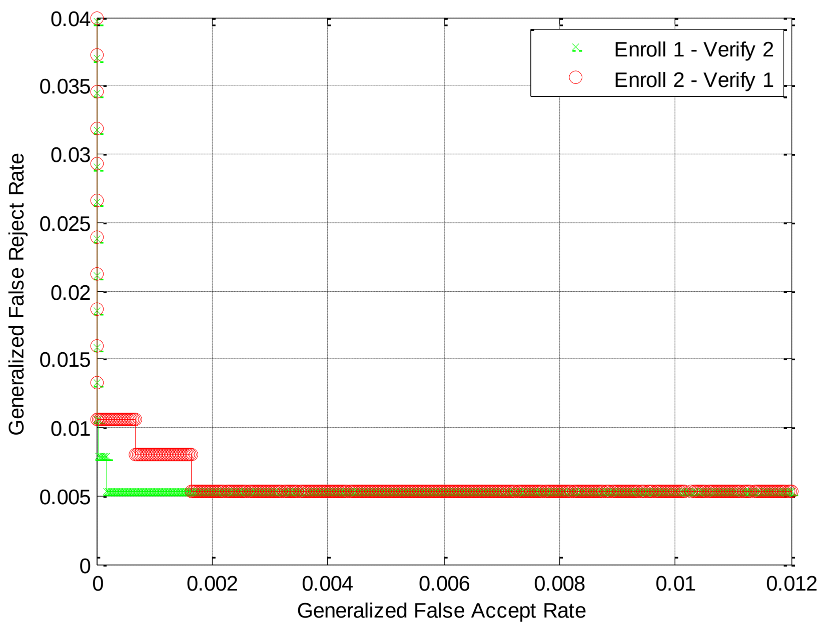


Figure 39. Enrol C Verify I – GFRR Versus GFAR

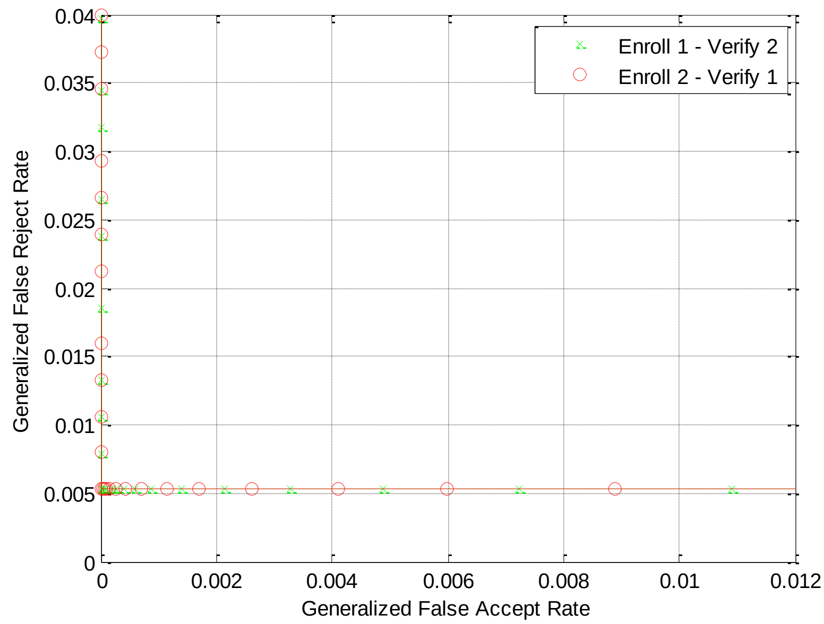


Figure 40. Enrol C Verify J – GFRR Versus GFAR

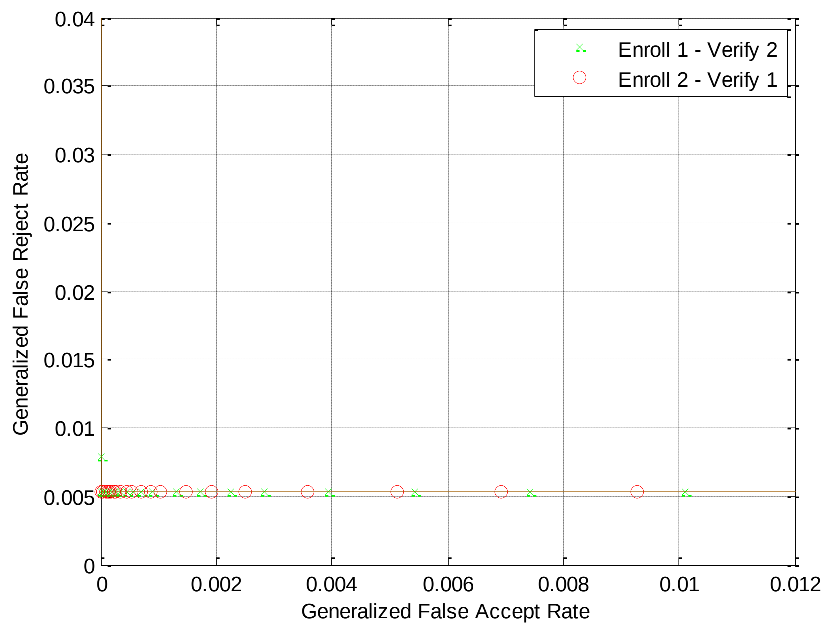


Figure 41. Enrol C Verify K – GFRR Versus GFAR

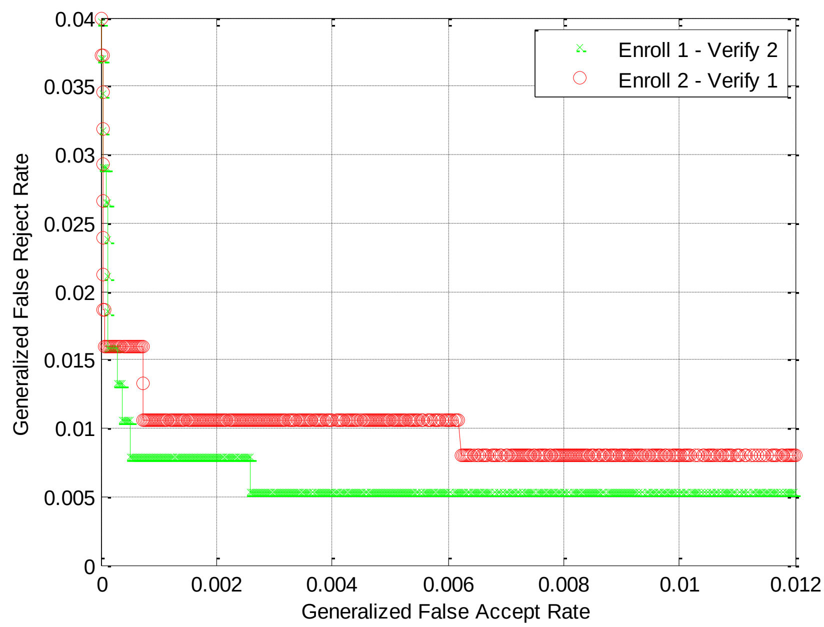


Figure 42. Enrol C Verify L – GFRR Versus GFAR

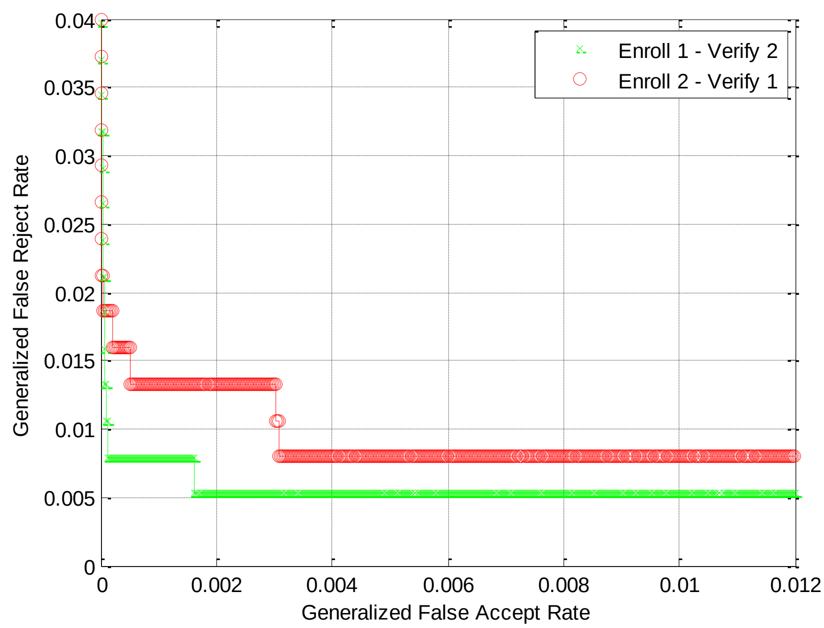


Figure 43. Enrol D Verify A – GFRR Versus GFAR

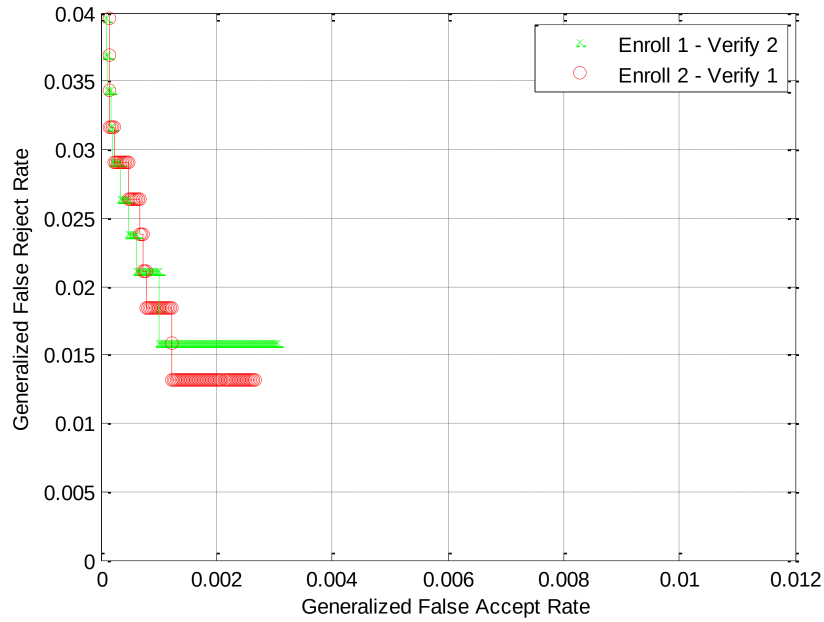


Figure 44. Enrol D Verify B – GFRR Versus GFAR

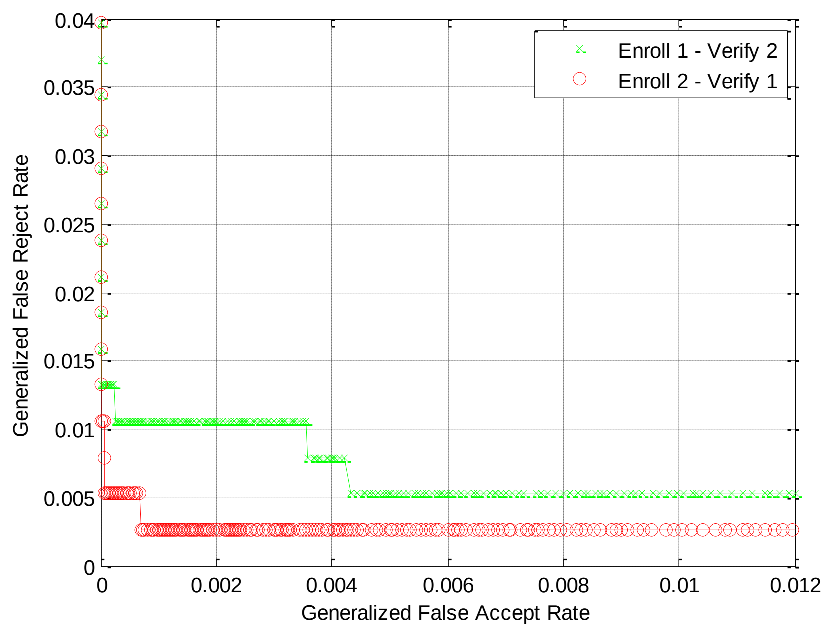


Figure 45. Enrol D Verify C – GFRR Versus GFAR

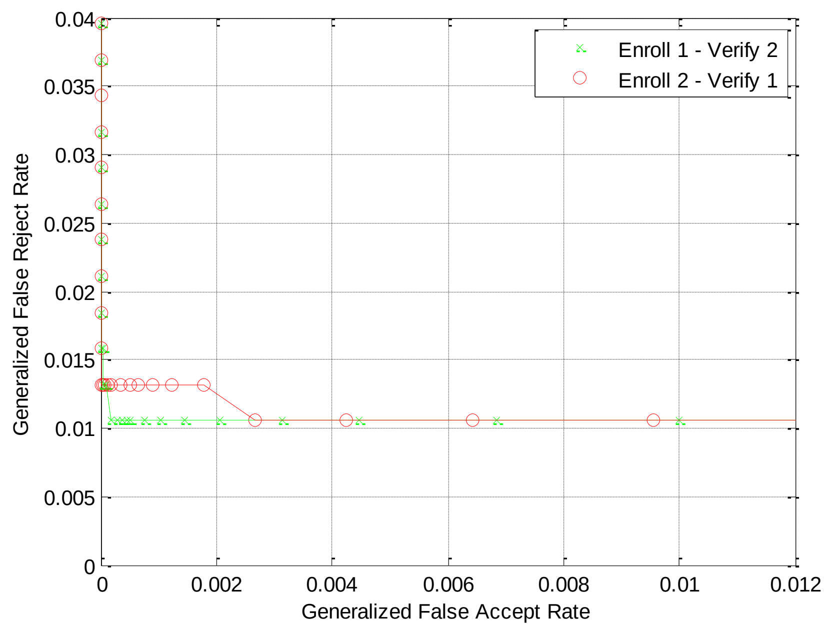


Figure 46. Enrol D Verify D – GFRR Versus GFAR

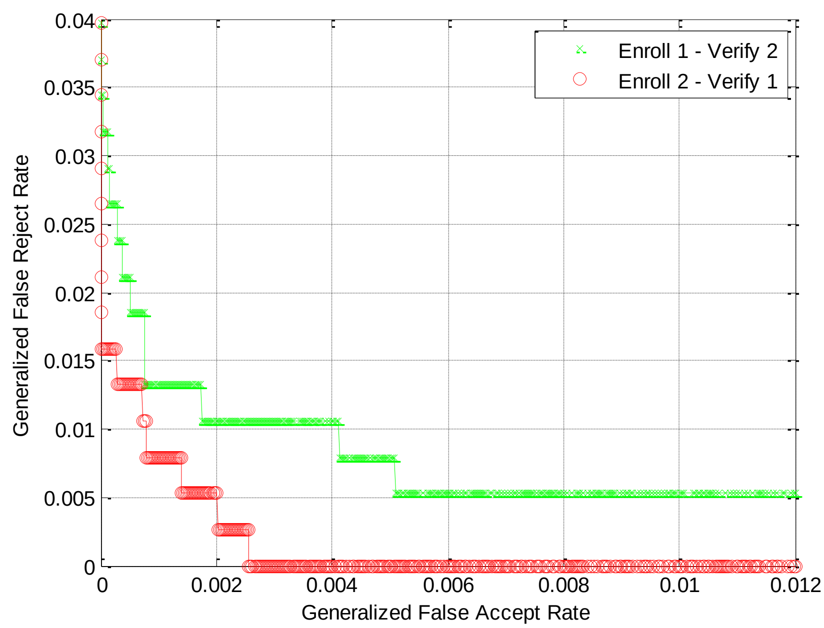


Figure 47. Enrol D Verify E – GFRR Versus GFAR

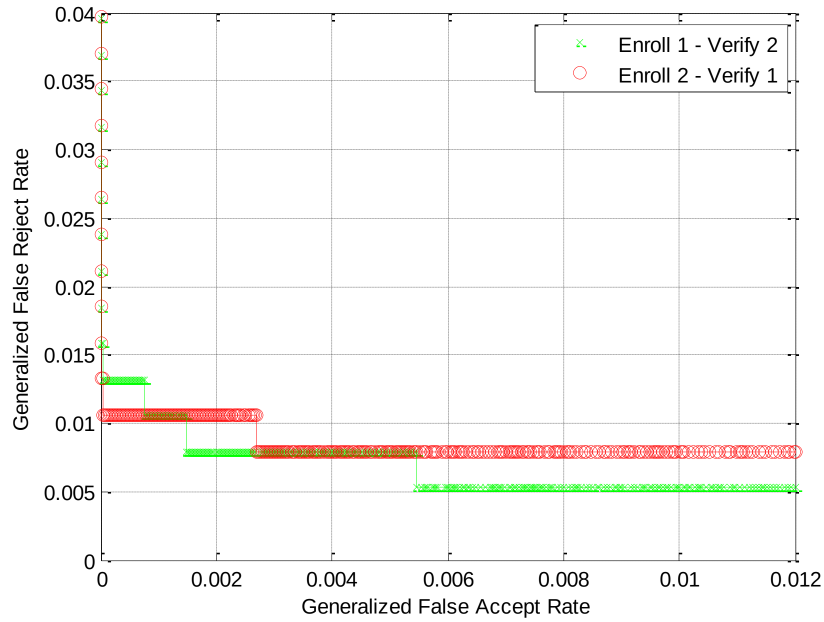


Figure 48. Enrol D Verify F – GFRR Versus GFAR

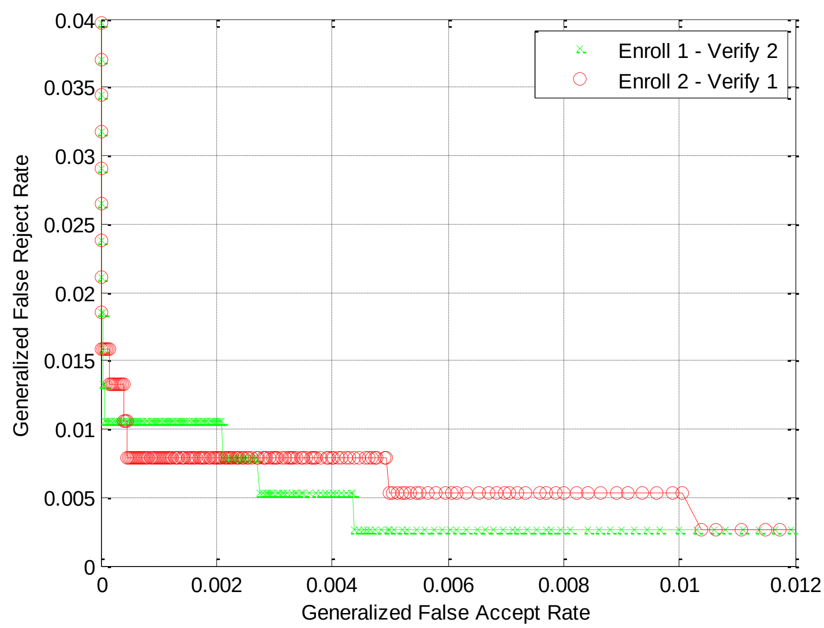


Figure 49. Enrol D Verify G – GFRR Versus GFAR

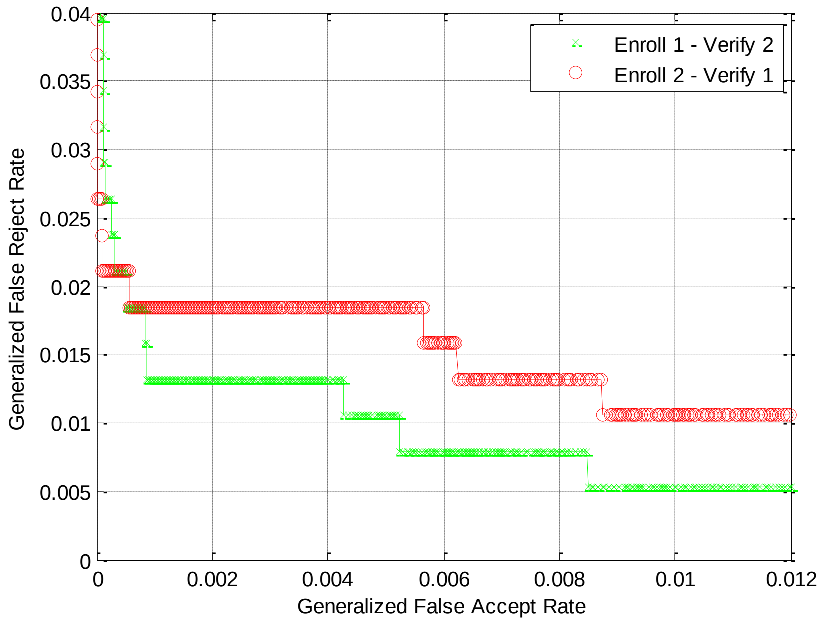


Figure 50. Enrol D Verify H – GFRR Versus GFAR

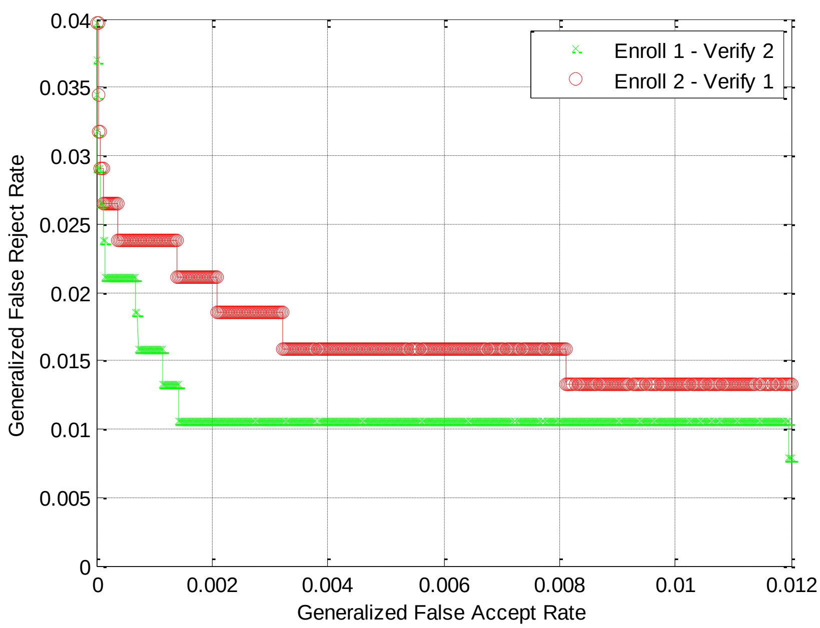


Figure 51. Enrol D Verify I – GFRR Versus GFAR

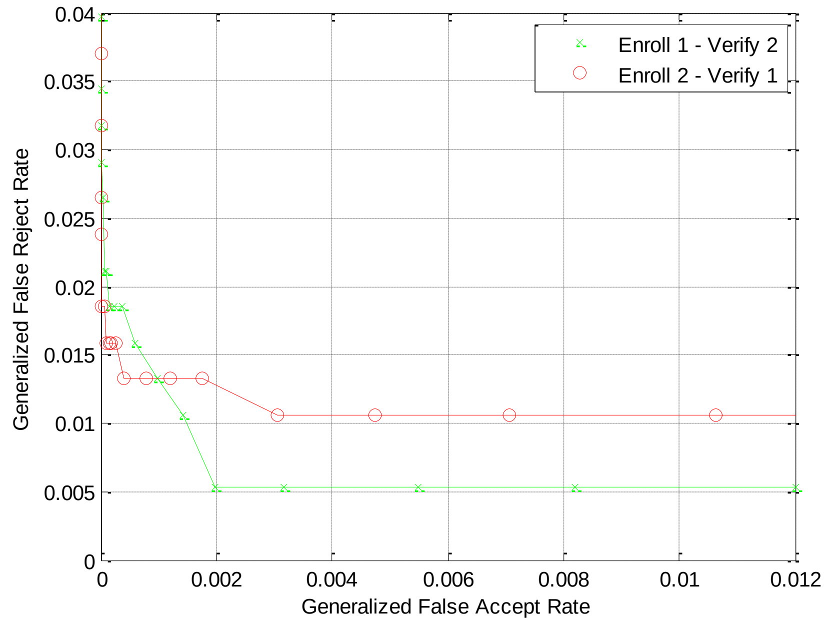


Figure 52. Enrol D Verify J – GFRR Versus GFAR

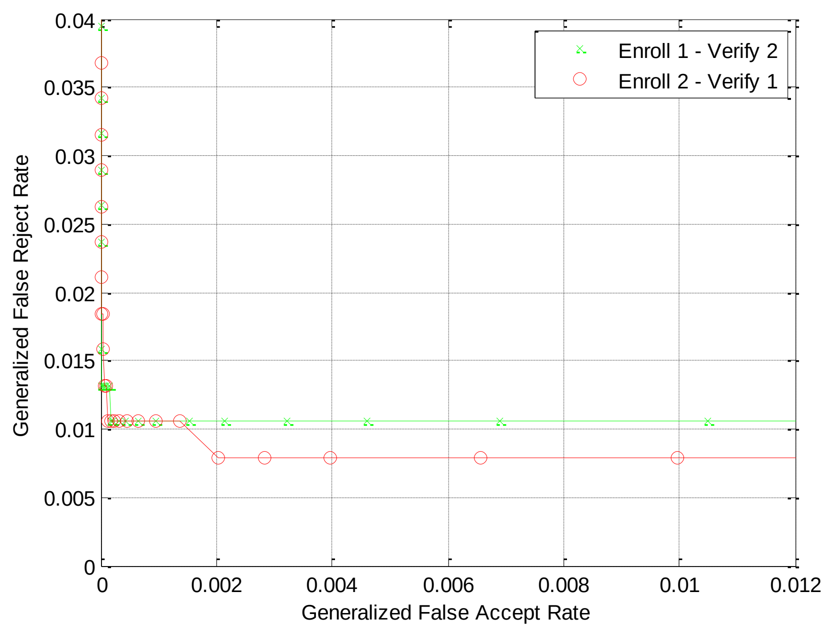


Figure 53. Enrol D Verify K – GFRR Versus GFAR

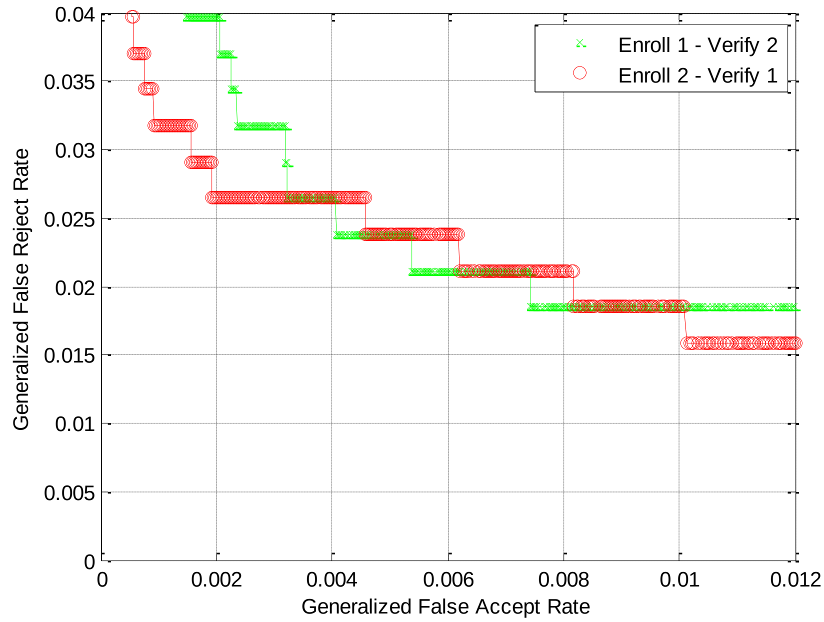


Figure 54. Enrol D Verify L – GFRR Versus GFAR

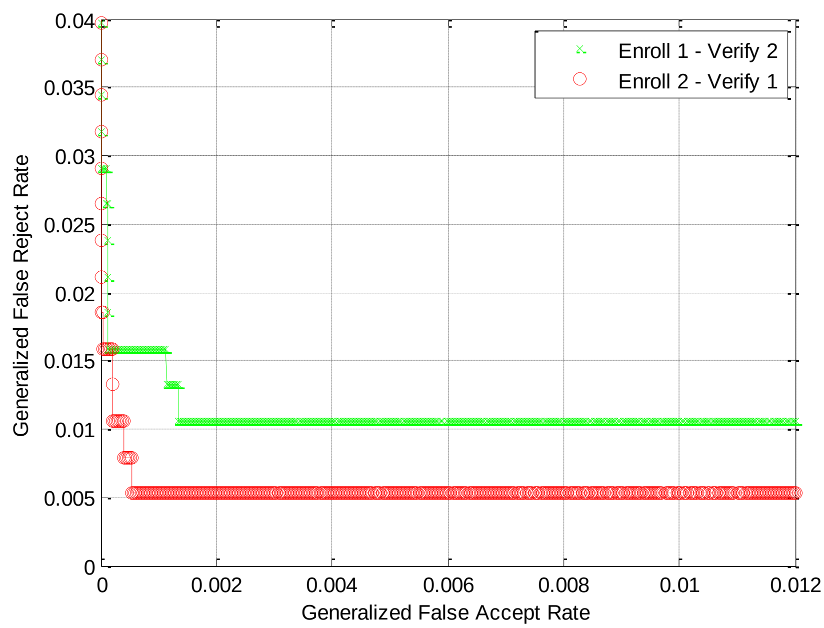


Figure 55. Enrol E Verify A – GFRR Versus GFAR

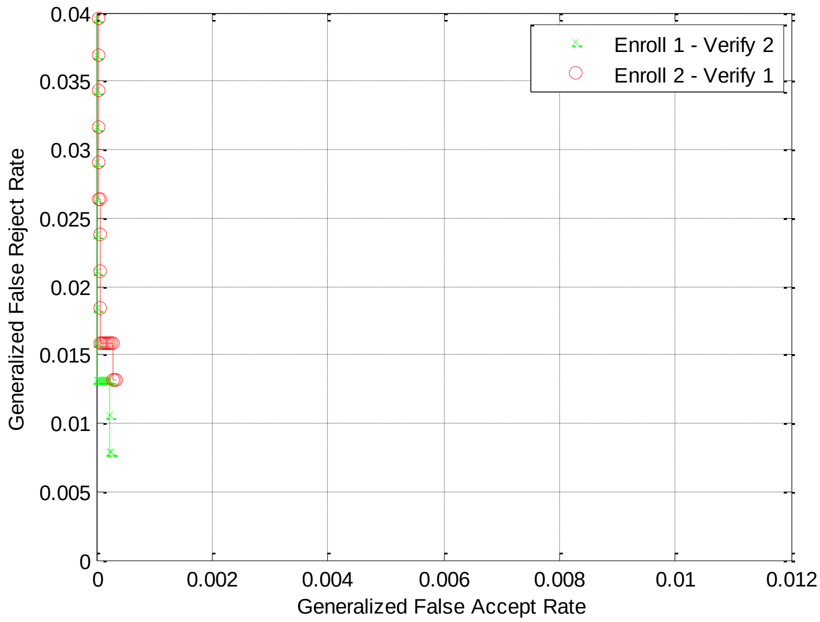


Figure 56. Enrol E Verify B – GFRR Versus GFAR

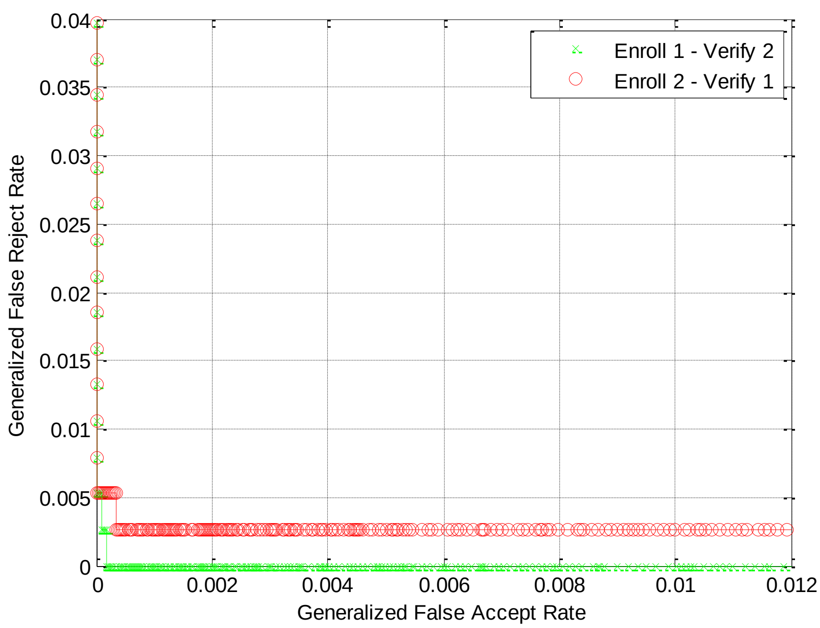


Figure 57. Enrol E Verify C – GFRR Versus GFAR

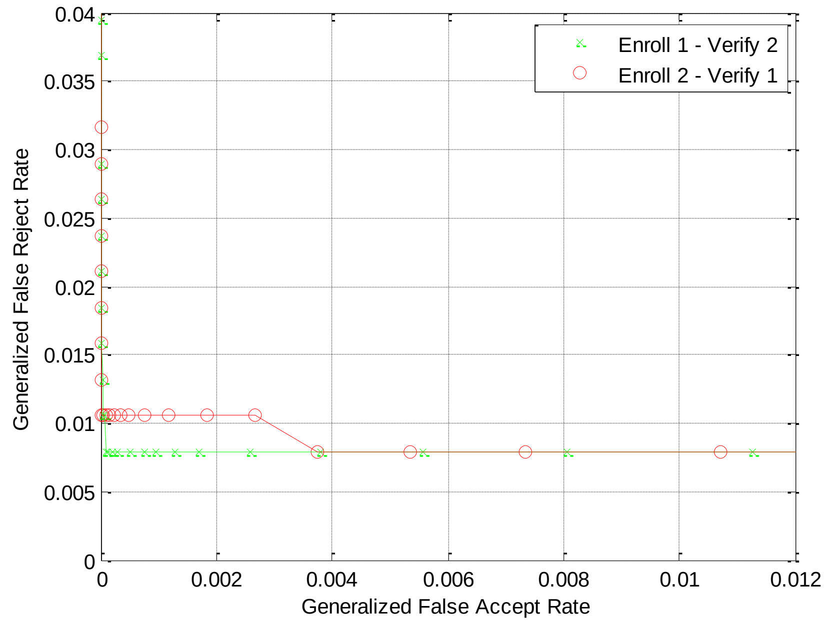


Figure 58. Enrol E Verify D – GFRR Versus GFAR

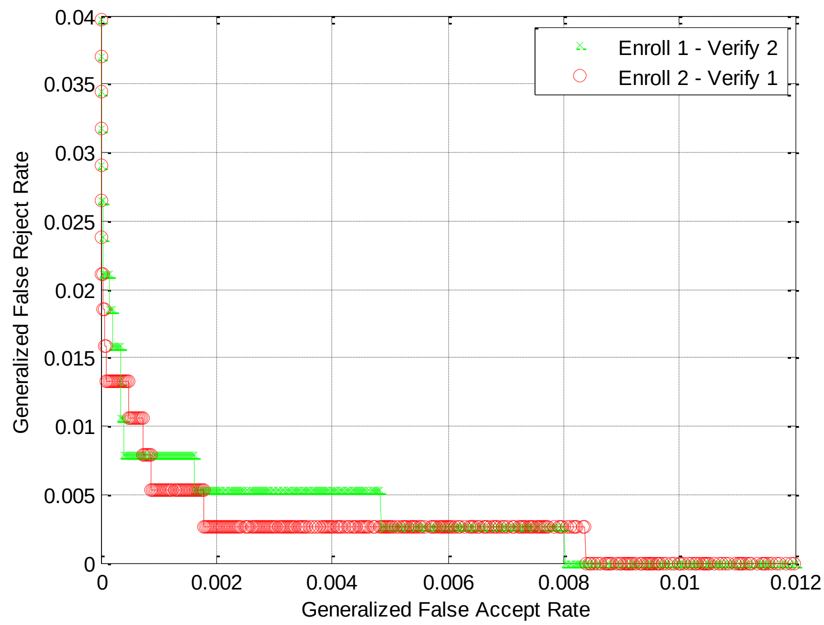


Figure 59. Enrol E Verify E – GFRR Versus GFAR

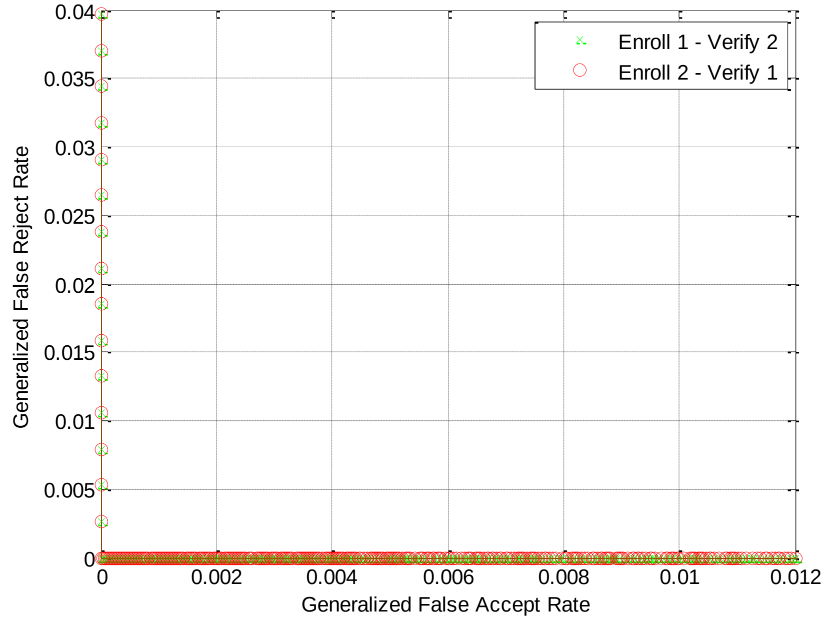


Figure 60. Enrol E Verify F – GFRR Versus GFAR

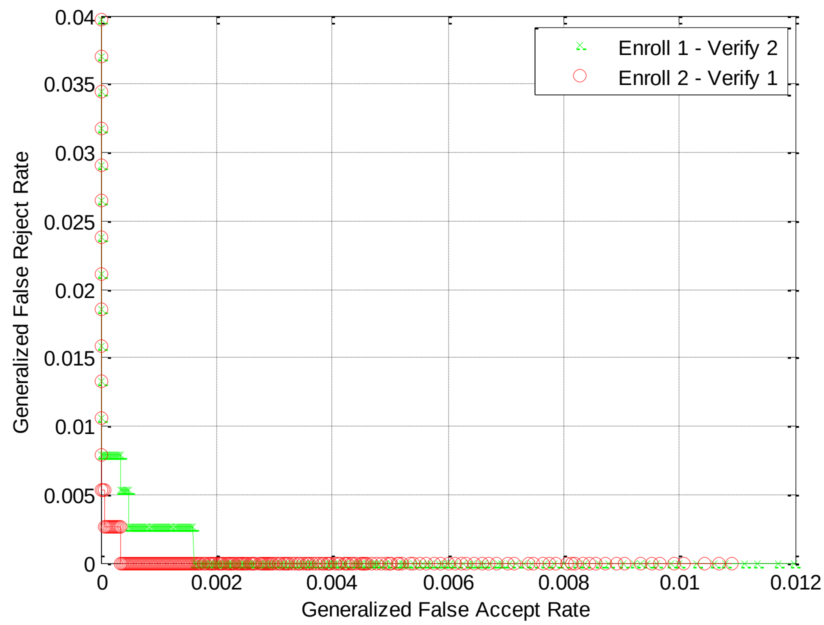


Figure 61. Enrol E Verify G – GFRR Versus GFAR

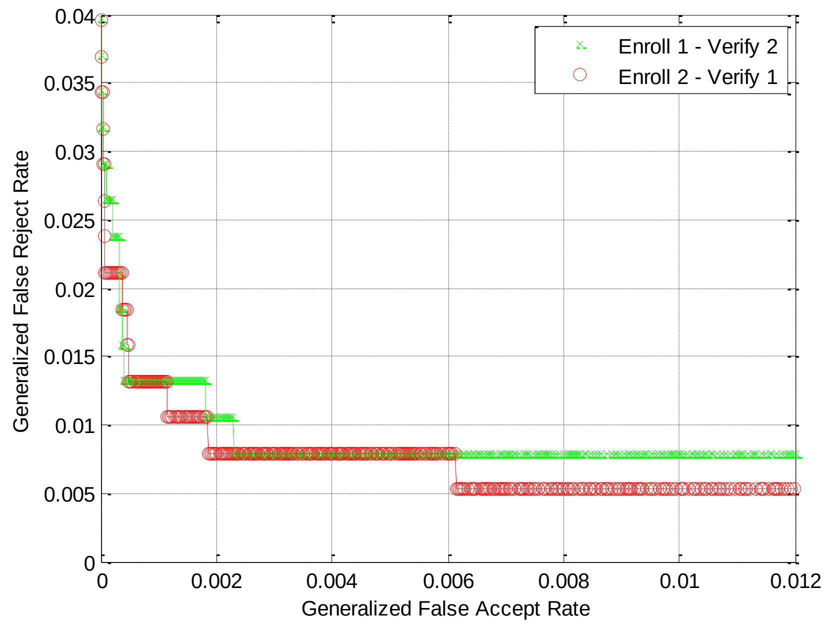


Figure 62. Enrol E Verify H – GFRR Versus GFAR

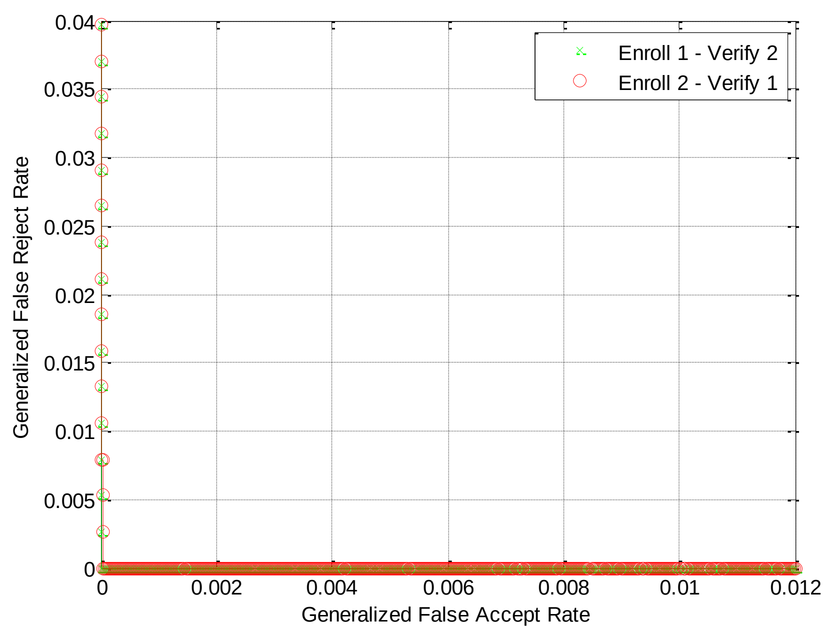


Figure 63. Enrol E Verify I – GFRR Versus GFAR

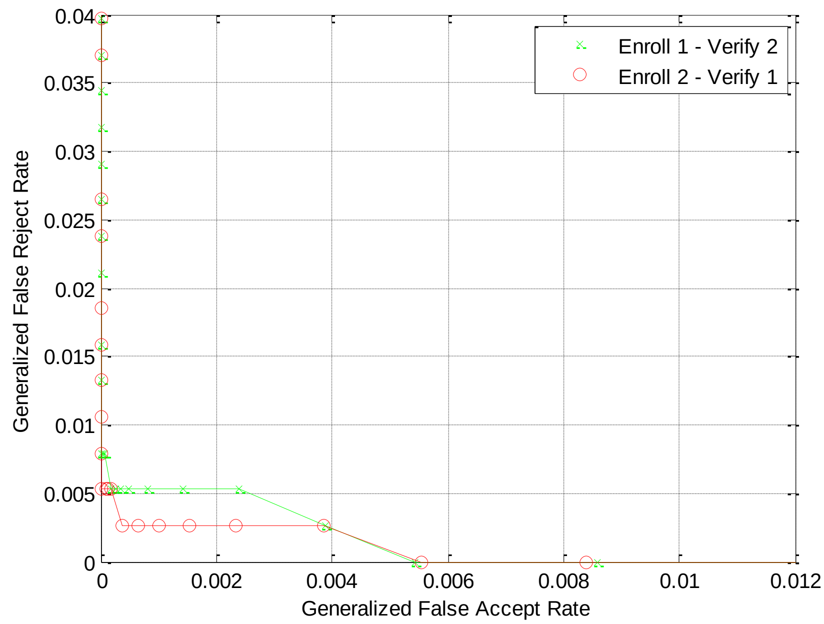


Figure 64. Enrol E Verify J – GFRR Versus GFAR

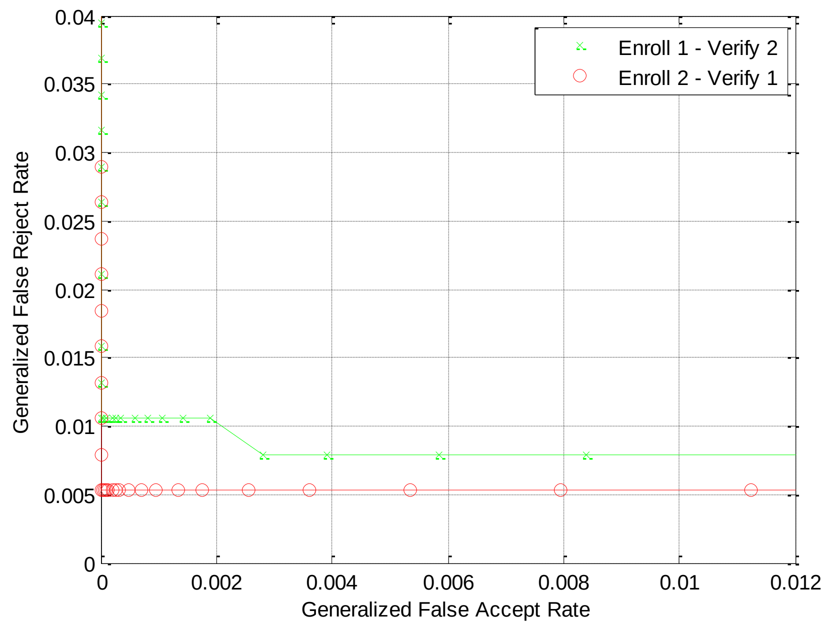


Figure 65. Enrol E Verify K – GFRR Versus GFAR

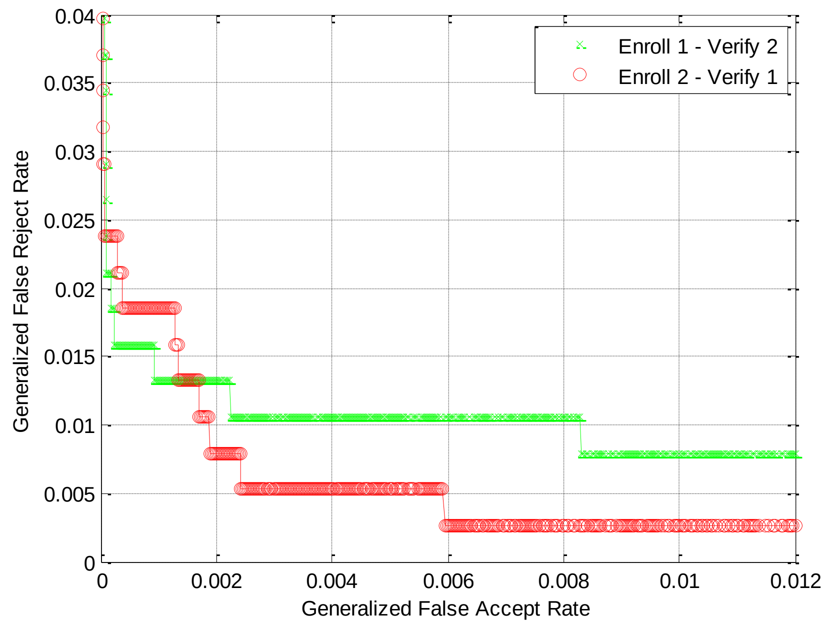


Figure 66. Enrol E Verify L – GFRR Versus GFAR

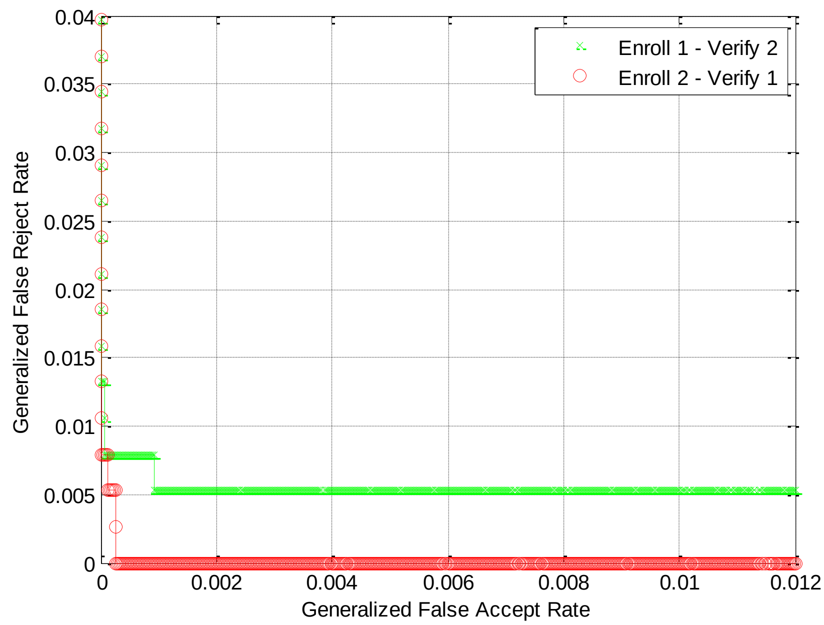


Figure 67. Enrol F Verify A – GFRR Versus GFAR

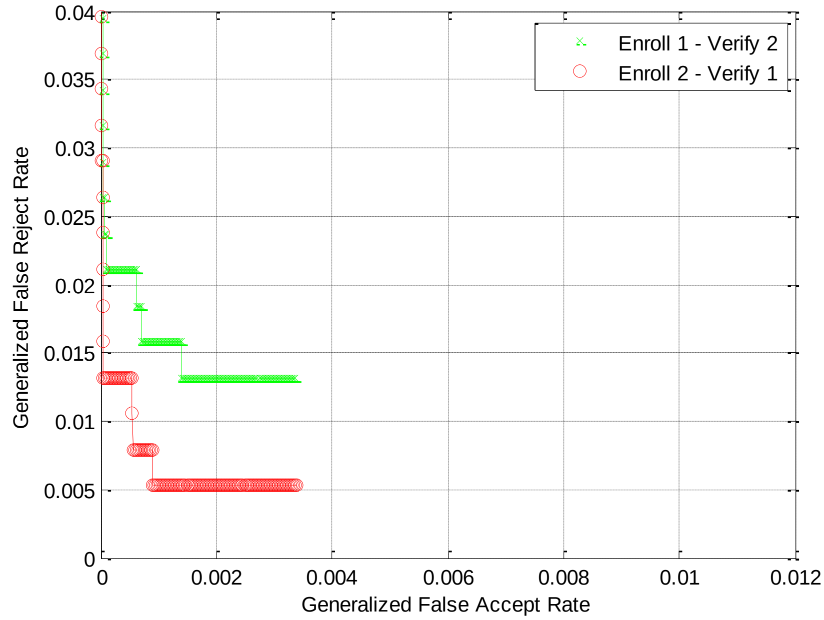


Figure 68. Enrol F Verify B – GFRR Versus GFAR

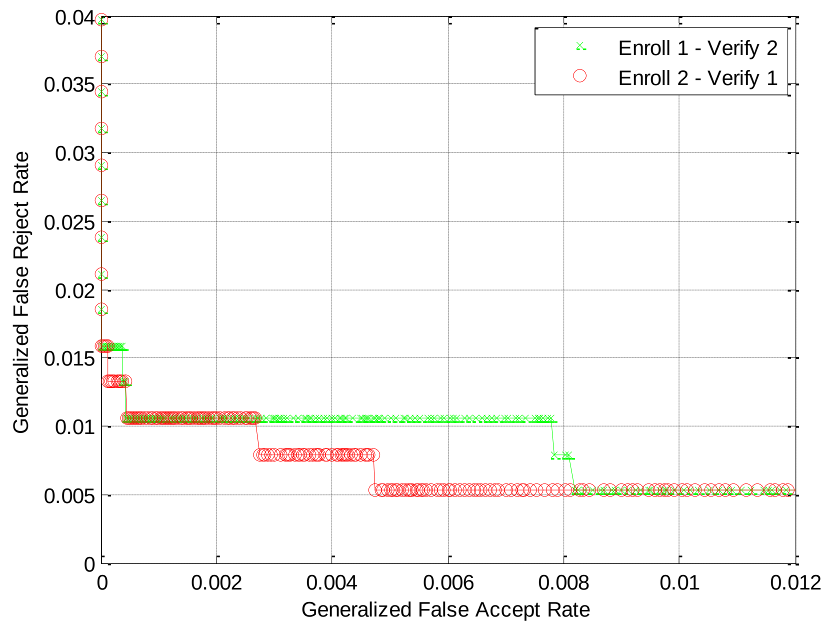


Figure 69. Enrol F Verify C – GFRR Versus GFAR

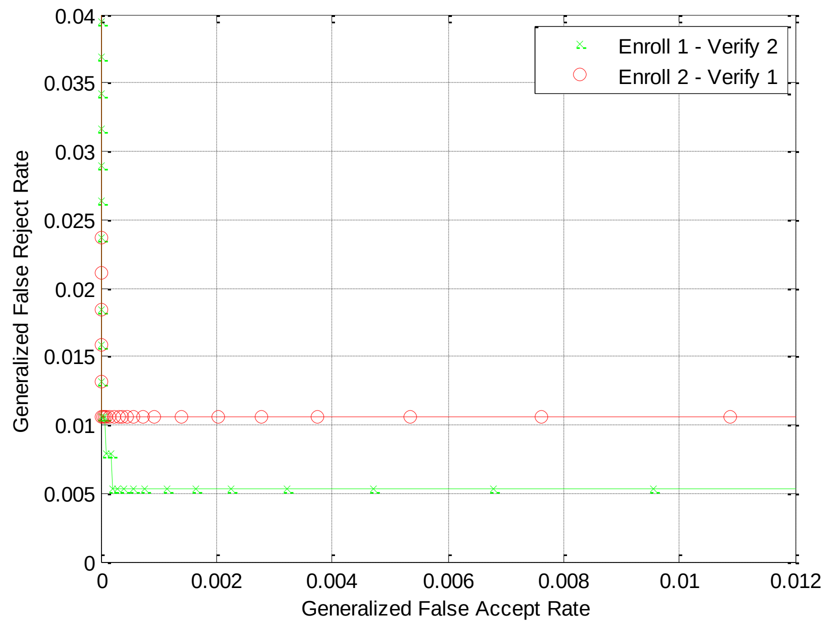


Figure 70. Enrol F Verify D – GFRR Versus GFAR

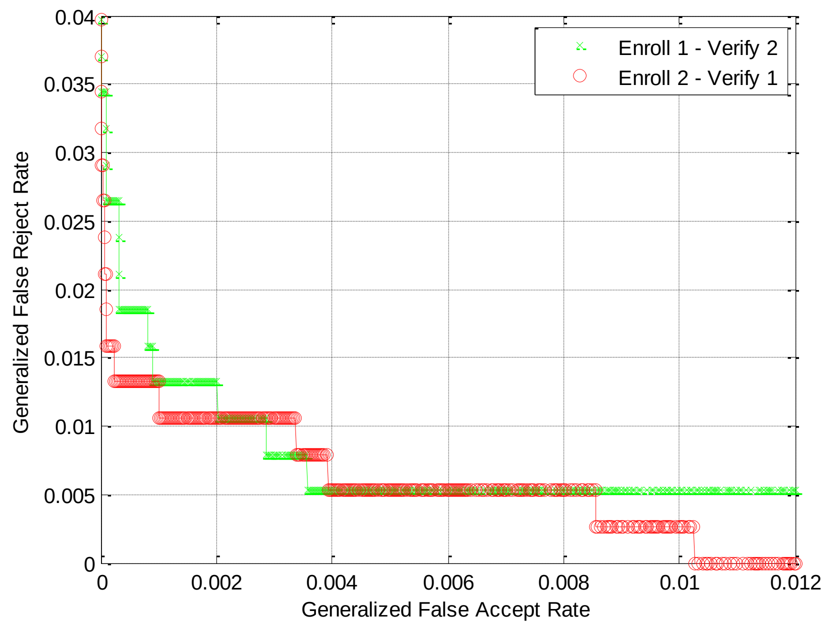


Figure 71. Enrol F Verify E – GFRR Versus GFAR

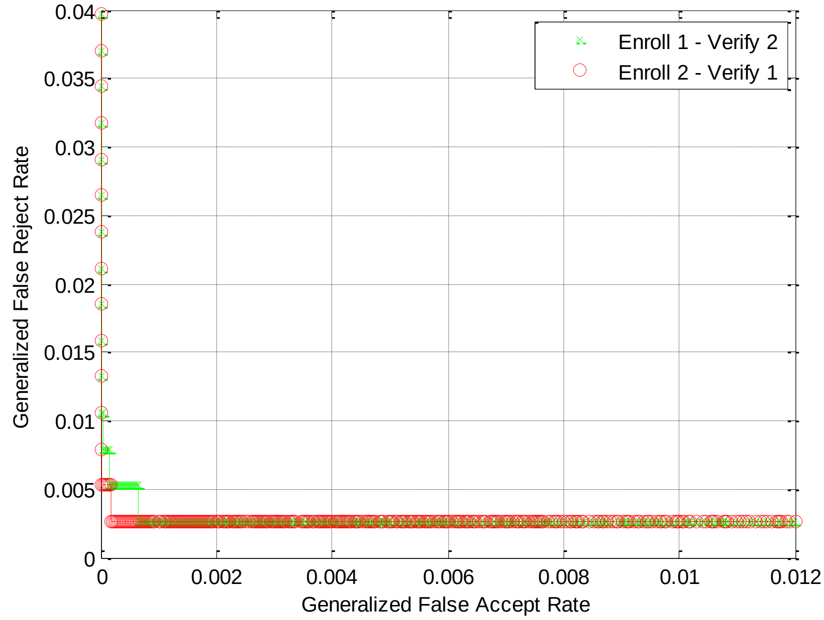


Figure 72. Enrol F Verify F – GFRR Versus GFAR

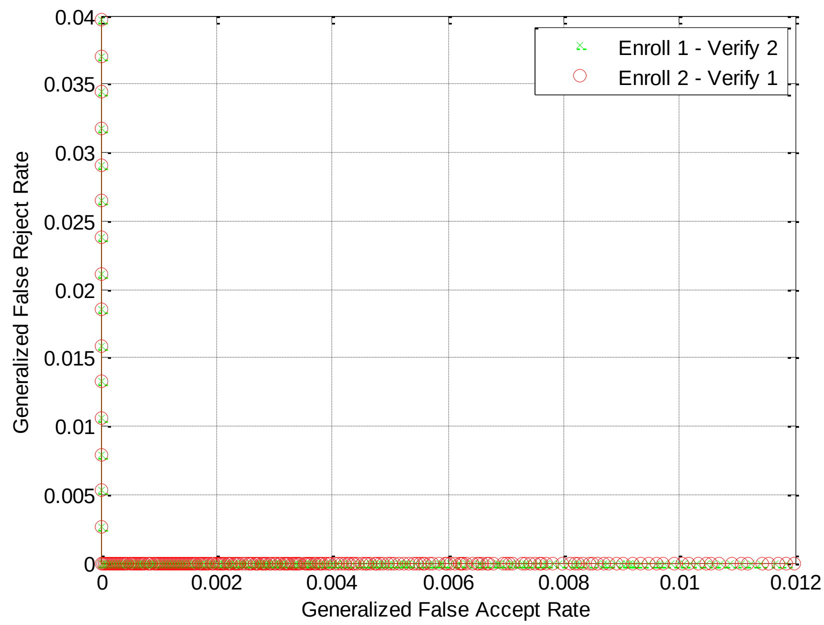


Figure 73. Enrol F Verify G – GFRR Versus GFAR

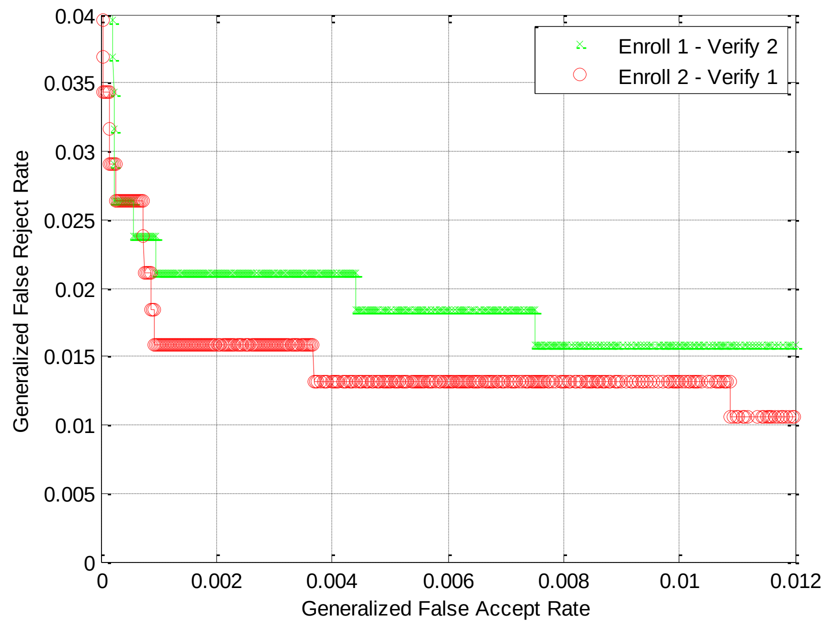


Figure 74. Enrol F Verify H – GFRR Versus GFAR

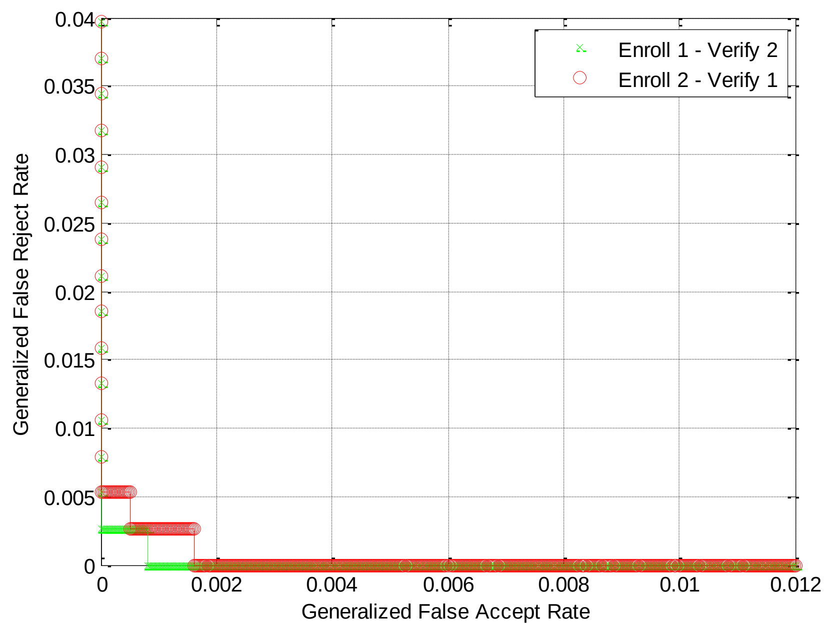


Figure 75. Enrol F Verify I – GFRR Versus GFAR

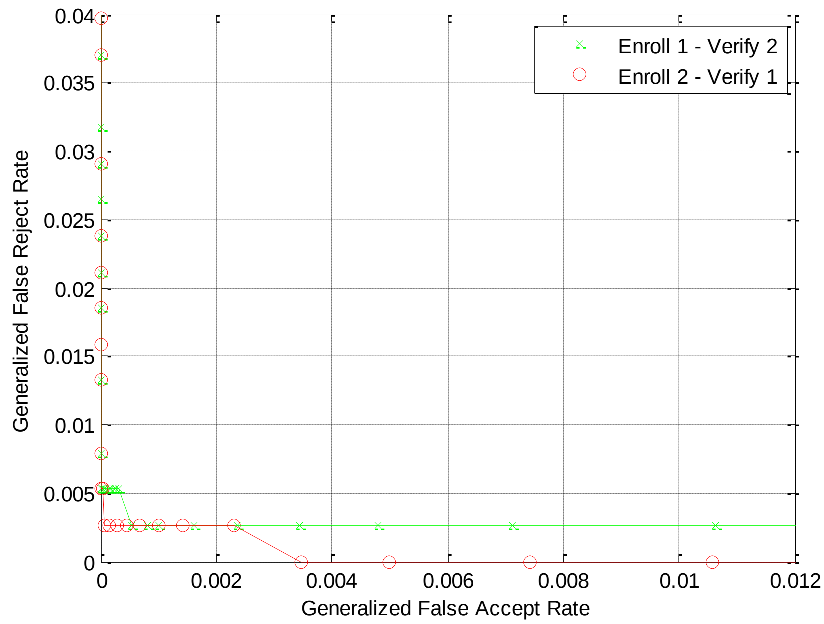


Figure 76. Enrol F Verify J – GFRR Versus GFAR

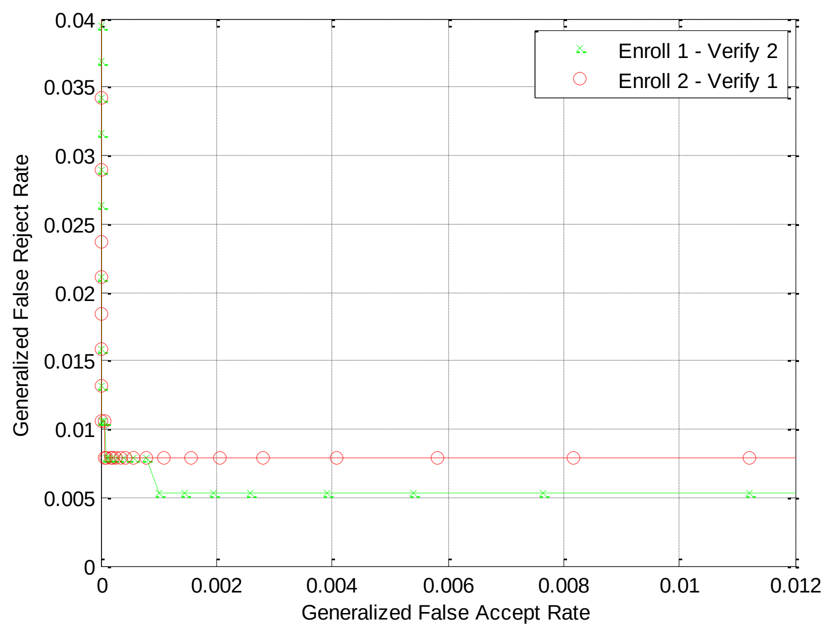


Figure 77. Enrol F Verify K – GFRR Versus GFAR

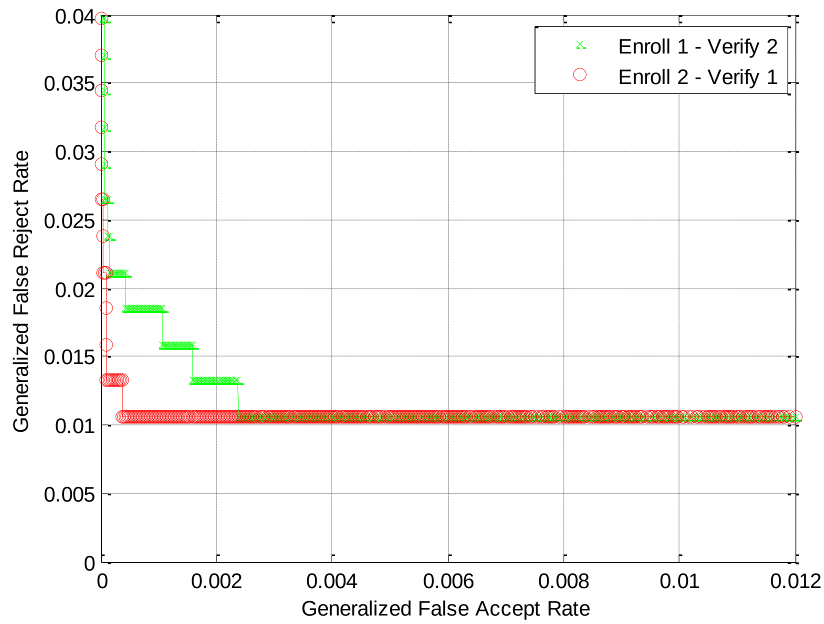


Figure 78. Enrol F Verify L – GFRR Versus GFAR

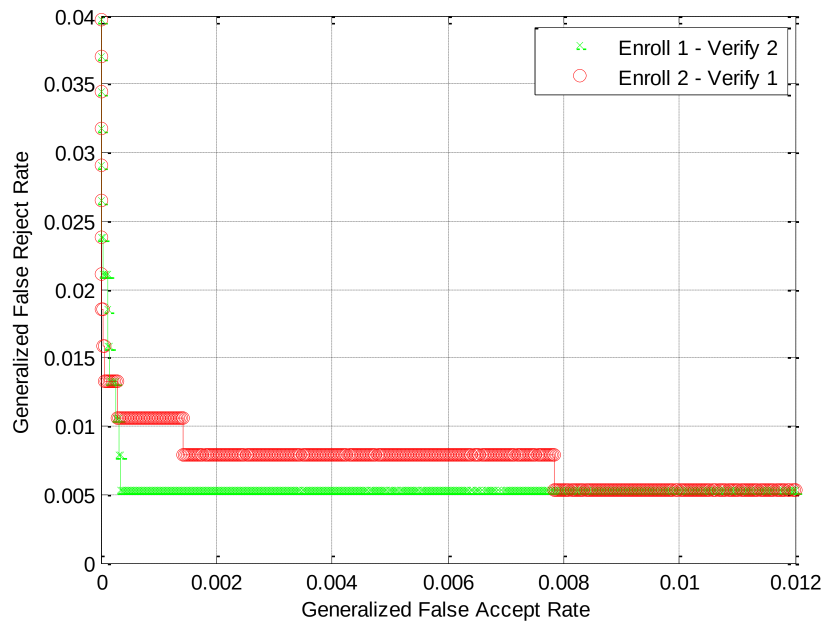


Figure 79. Enrol G Verify A – GFRR Versus GFAR

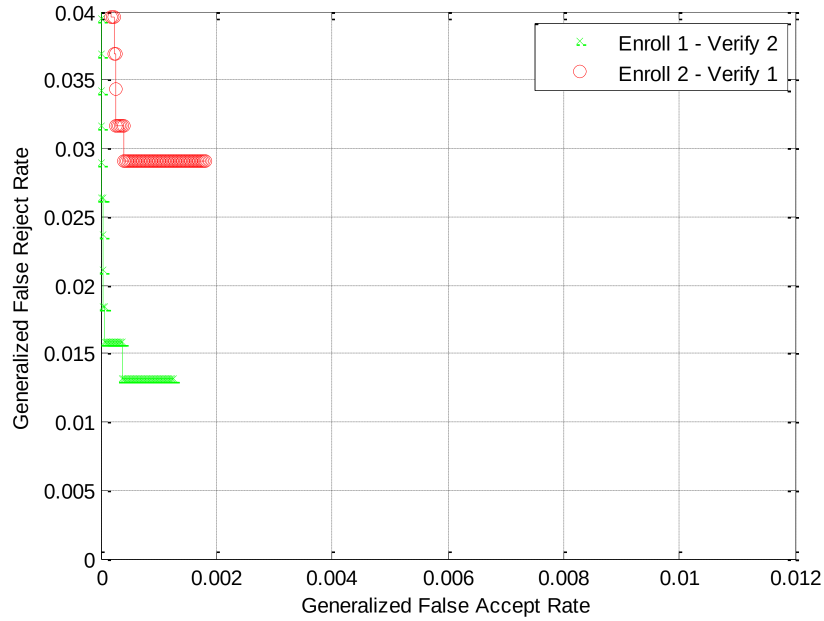


Figure 80. Enrol G Verify B – GFRR Versus GFAR

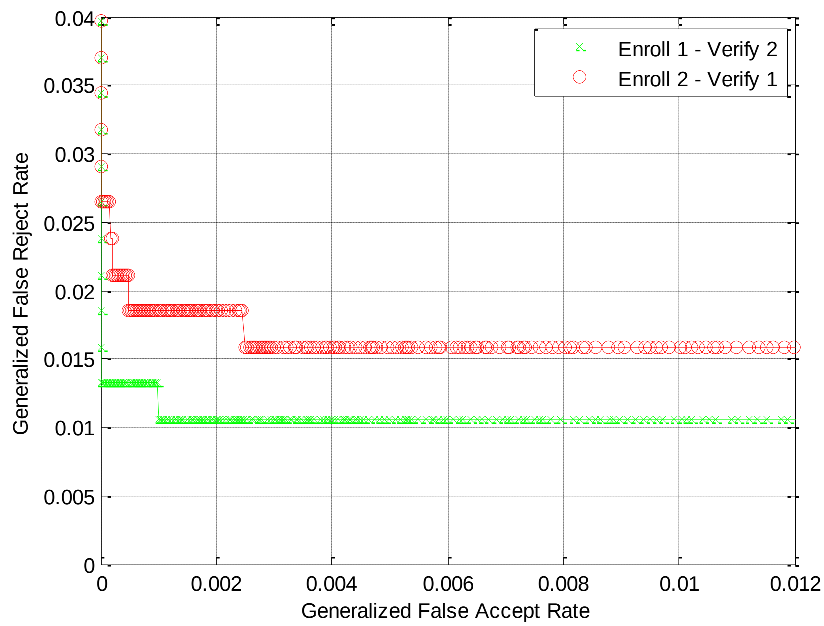


Figure 81. Enrol G Verify C – GFRR Versus GFAR

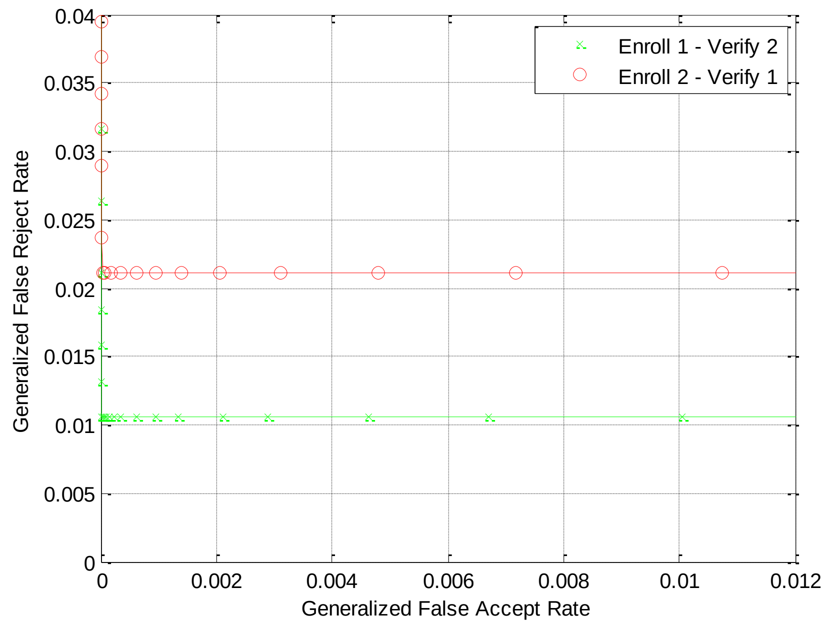


Figure 82. Enrol G Verify D – GFRR Versus GFAR

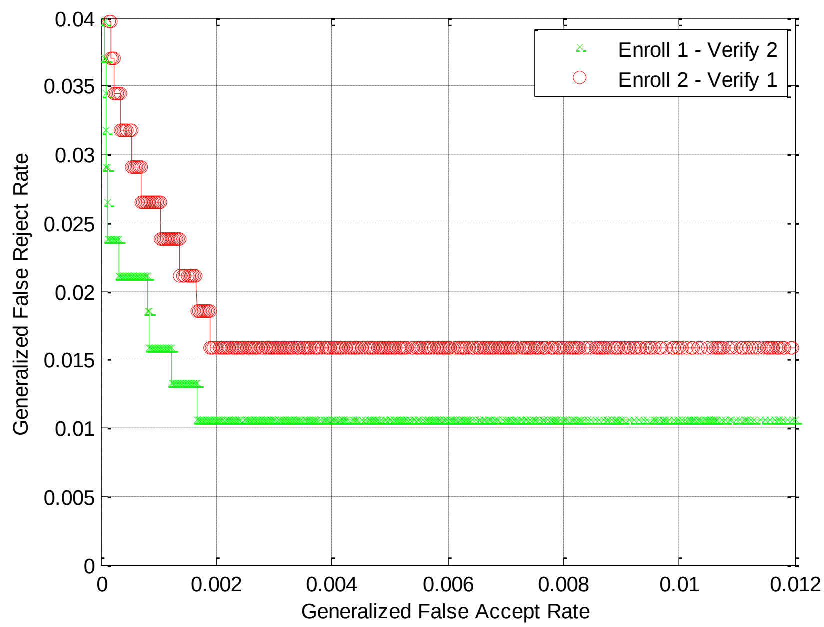


Figure 83. Enrol G Verify E – GFRR Versus GFAR

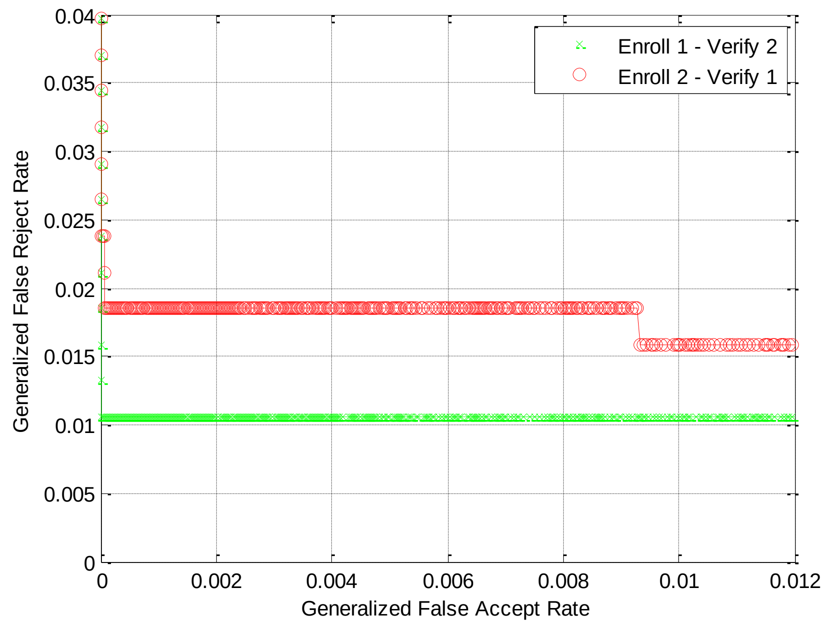


Figure 84. Enrol G Verify F – GFRR Versus GFAR

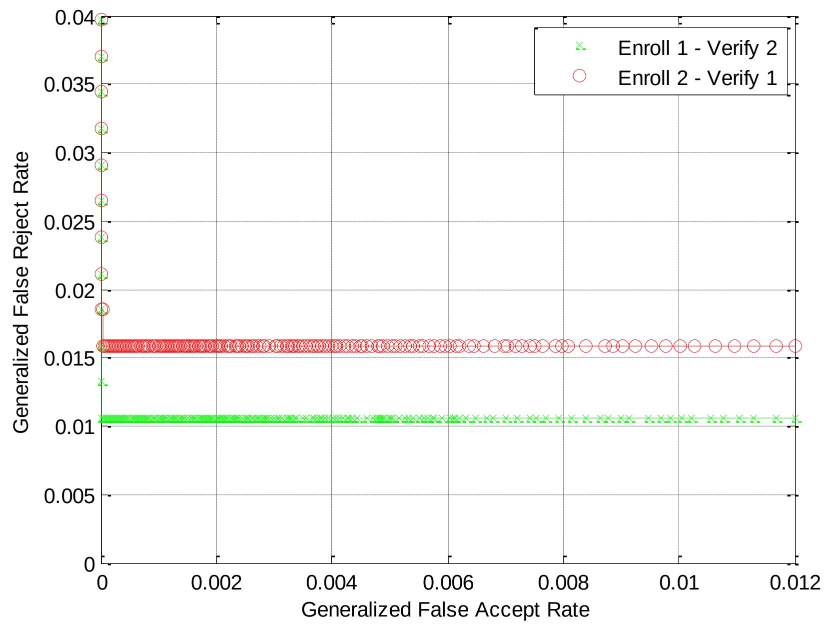


Figure 85. Enrol G Verify G – GFRR Versus GFAR

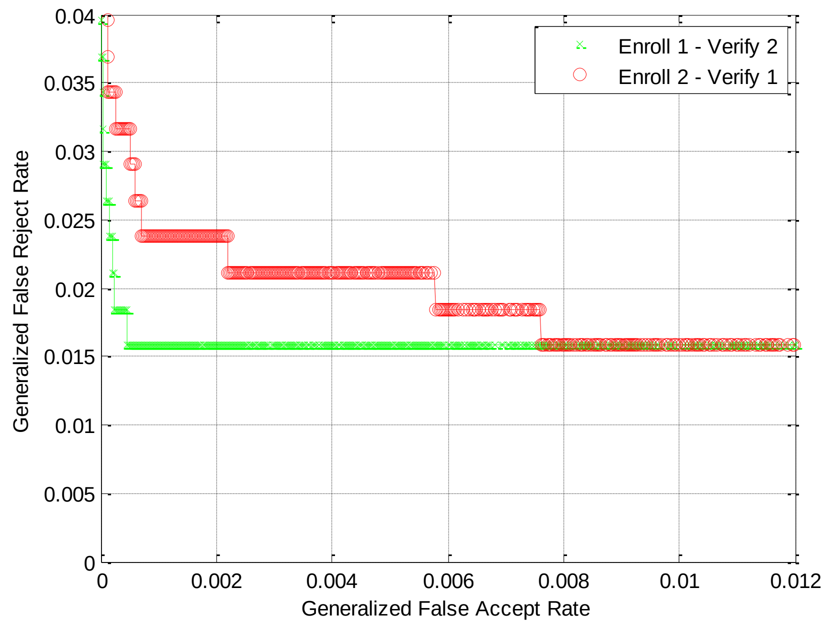


Figure 86. Enrol G Verify H – GFRR Versus GFAR

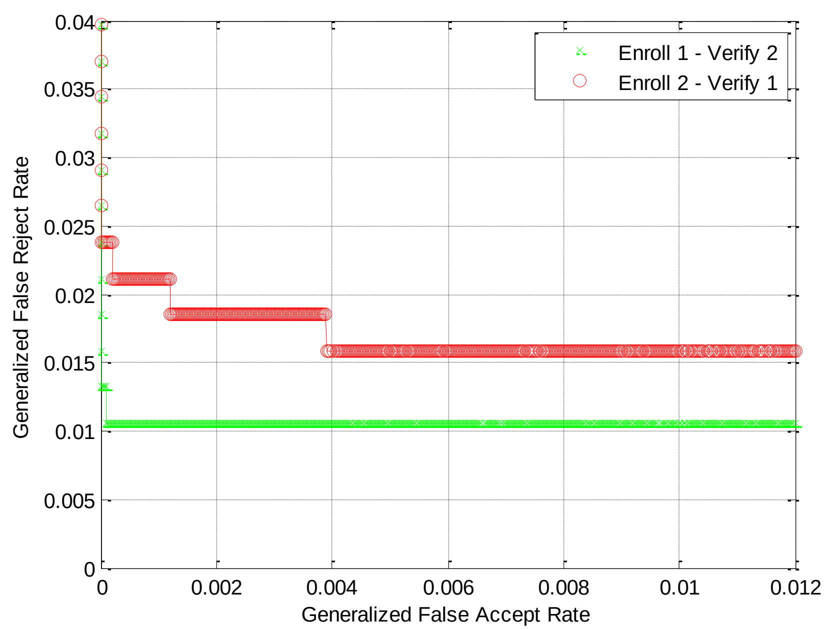


Figure 87. Enrol G Verify I – GFRR Versus GFAR

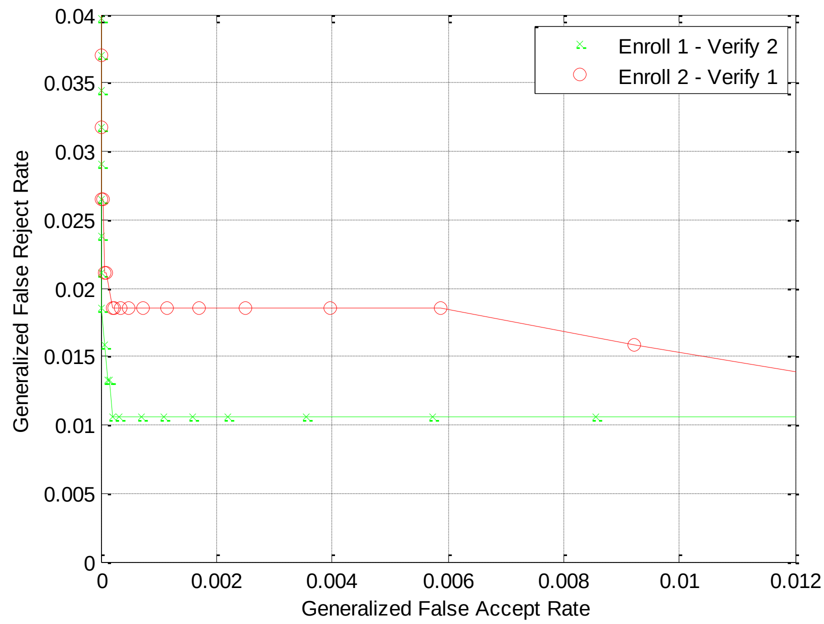


Figure 88. Enrol G Verify J – GFRR Versus GFAR

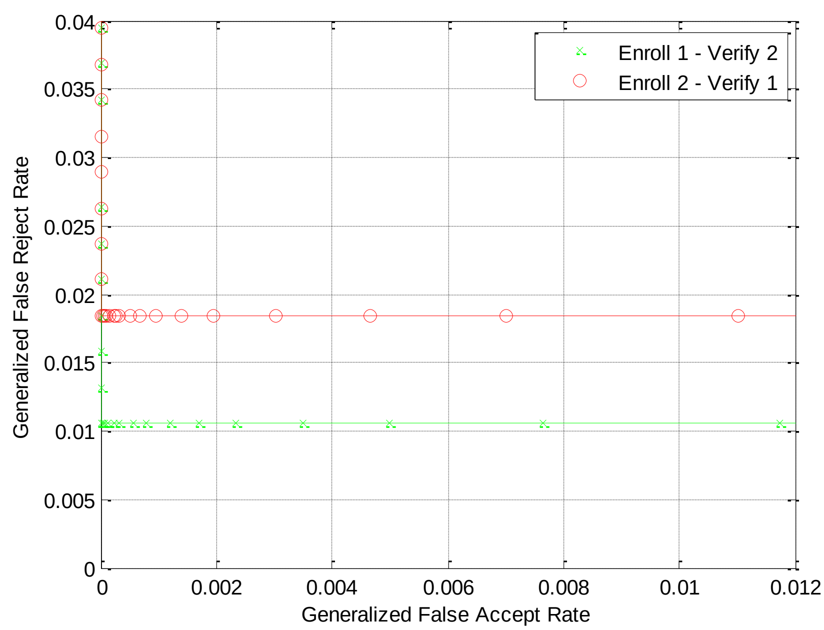


Figure 89. Enrol G Verify K – GFRR Versus GFAR

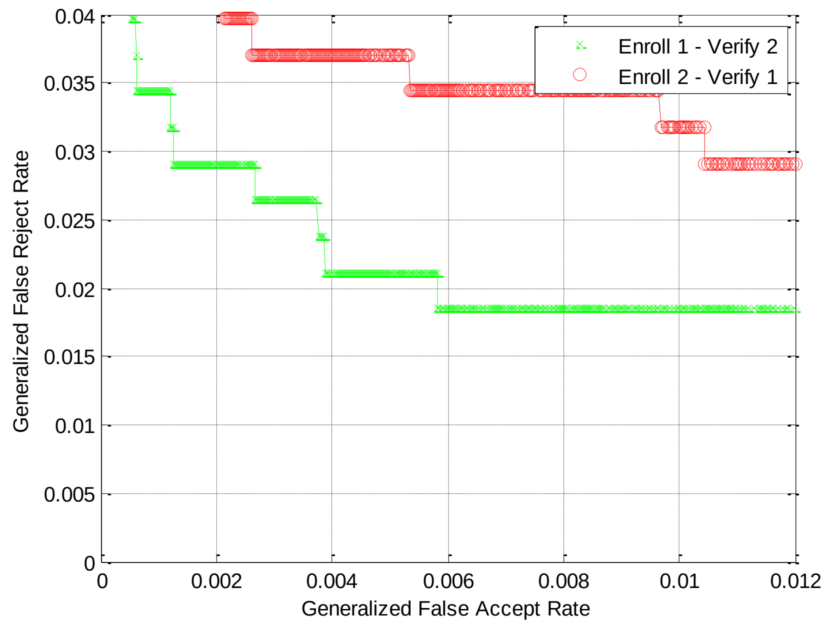


Figure 90. Enrol G Verify L – GFRR Versus GFAR

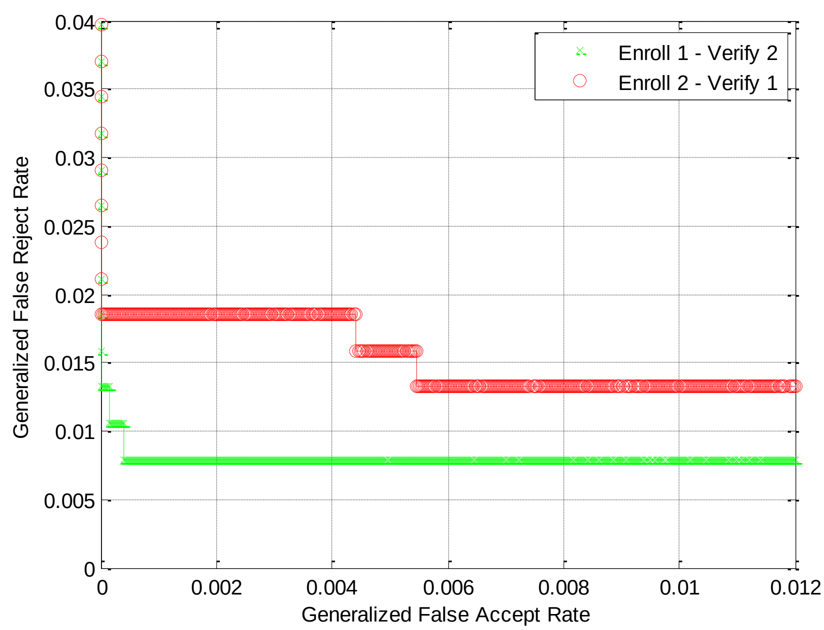


Figure 91. Enrol H Verify A – GFRR Versus GFAR

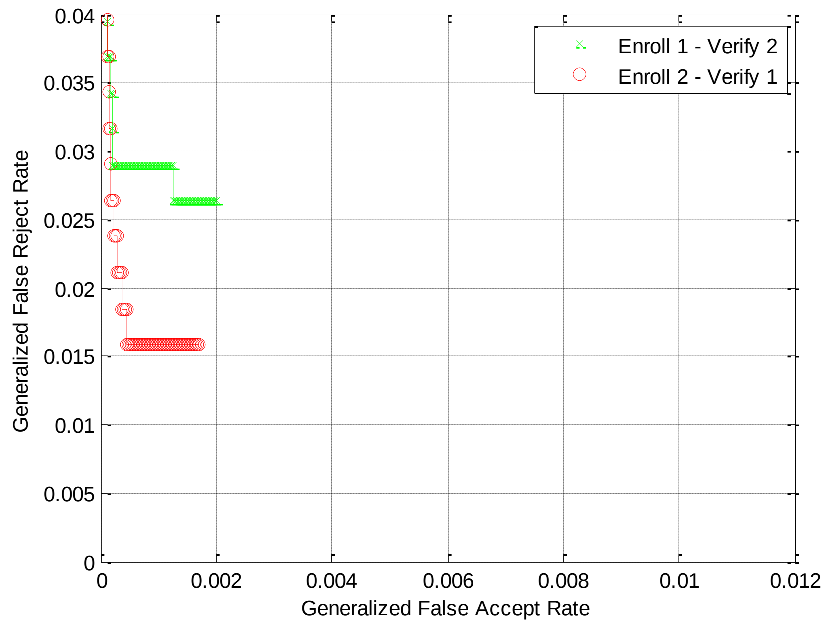


Figure 92. Enrol H Verify B – GFRR Versus GFAR

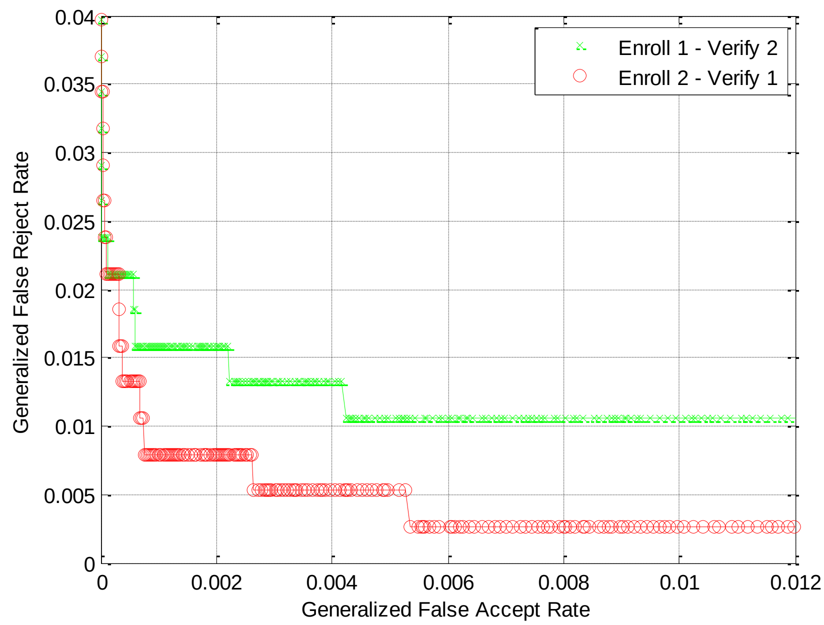


Figure 93. Enrol H Verify C – GFRR Versus GFAR

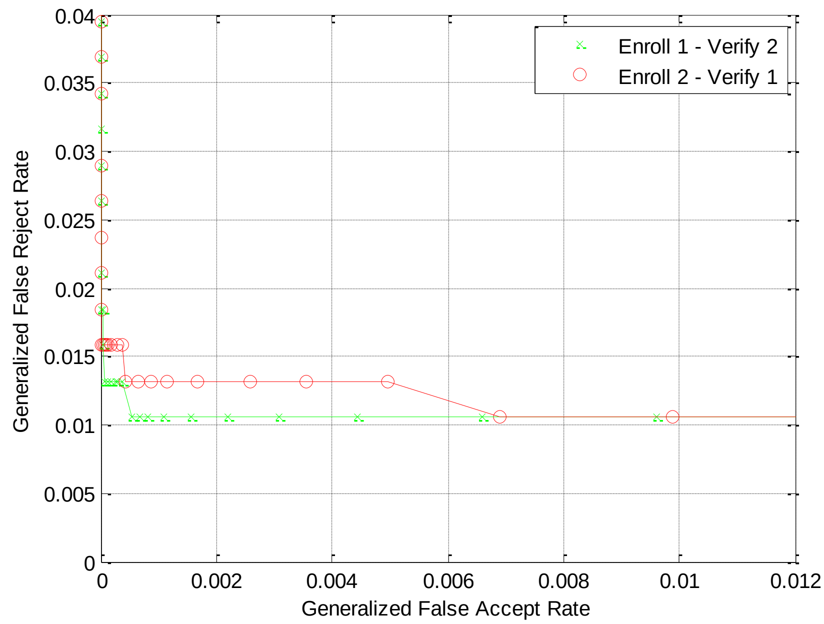


Figure 94. Enrol H Verify D – GFRR Versus GFAR

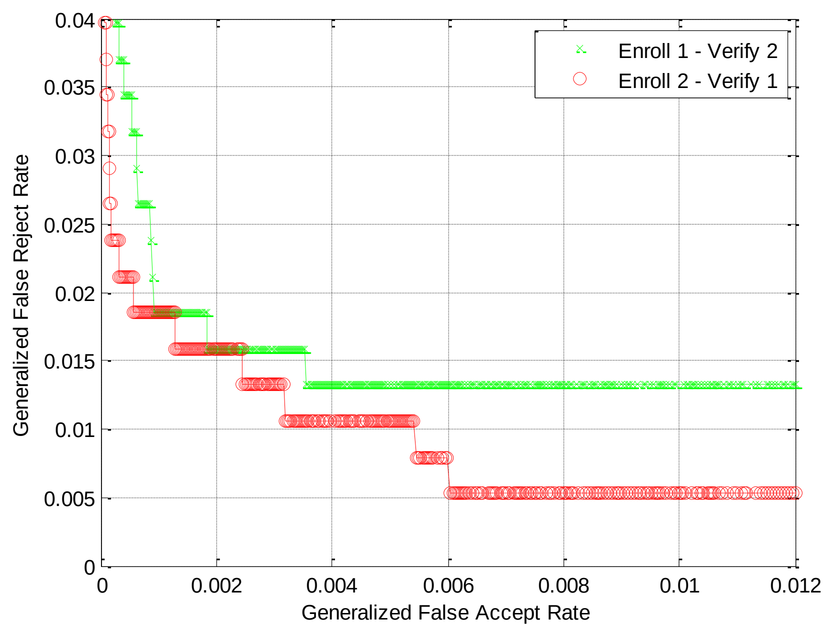


Figure 95. Enrol H Verify E – GFRR Versus GFAR

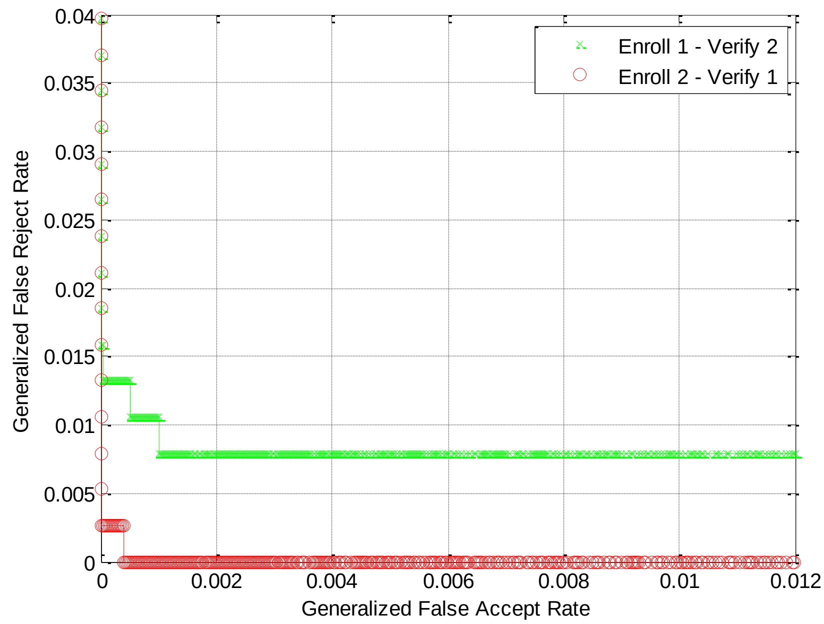


Figure 96. Enrol H Verify F – GFRR Versus GFAR

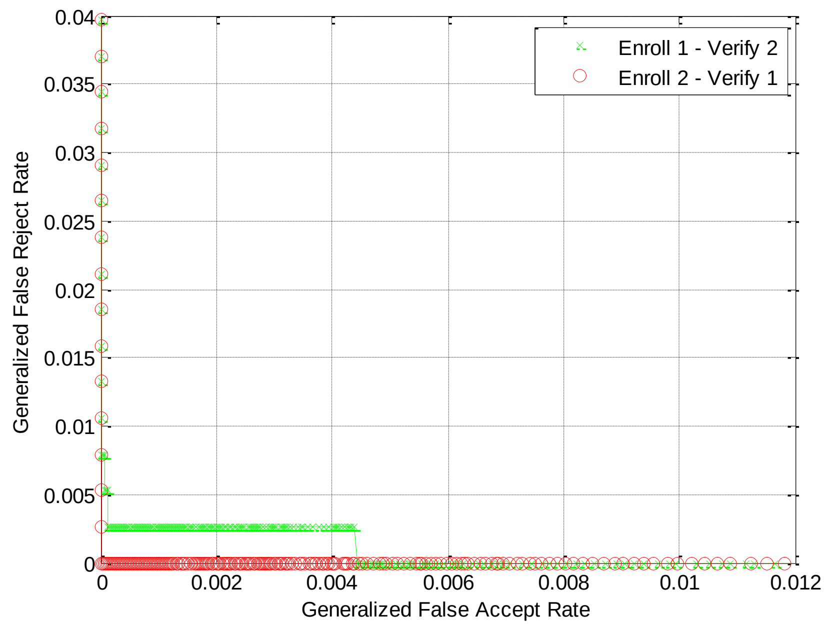


Figure 97. Enrol H Verify G – GFRR Versus GFAR

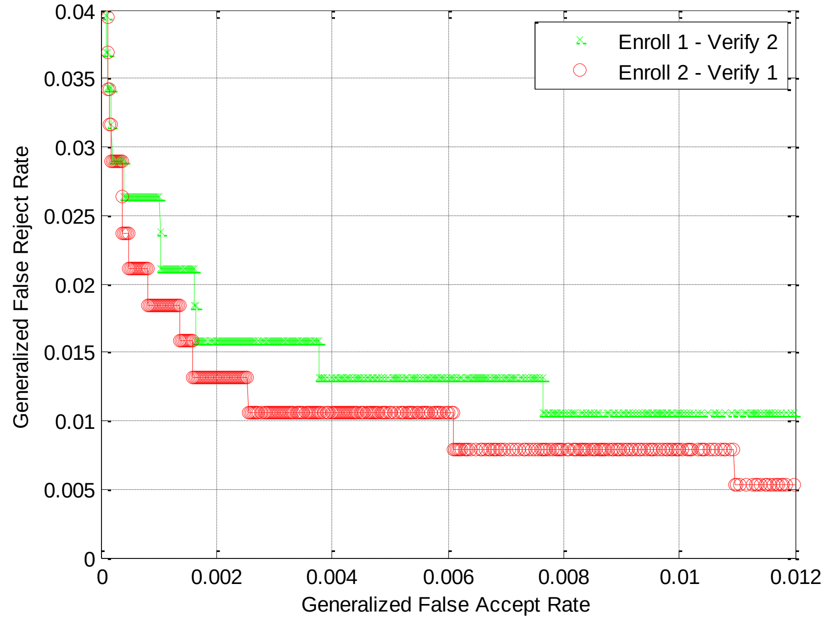


Figure 98. Enrol H Verify H – GFRR Versus GFAR

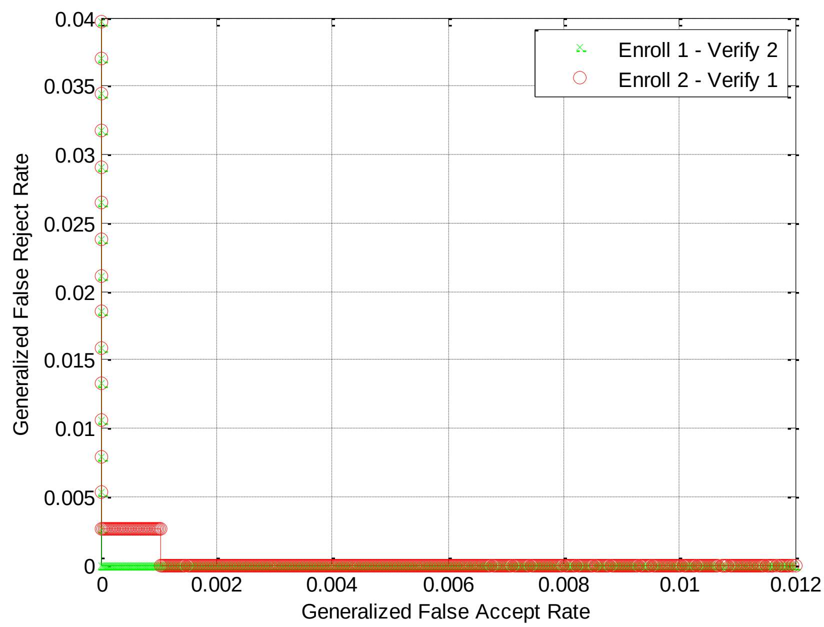


Figure 99. Enrol H Verify I – GFRR Versus GFAR

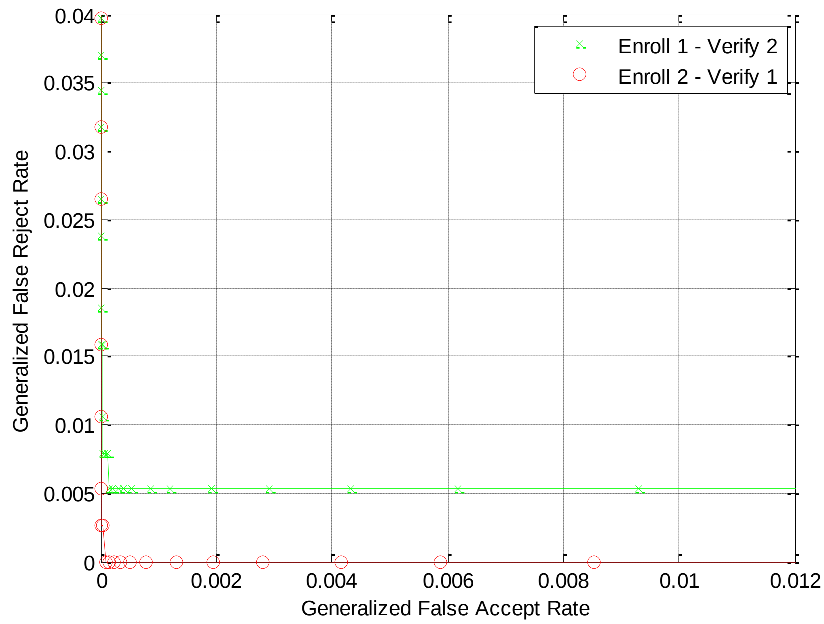


Figure 100. Enrol H Verify J – GFRR Versus GFAR

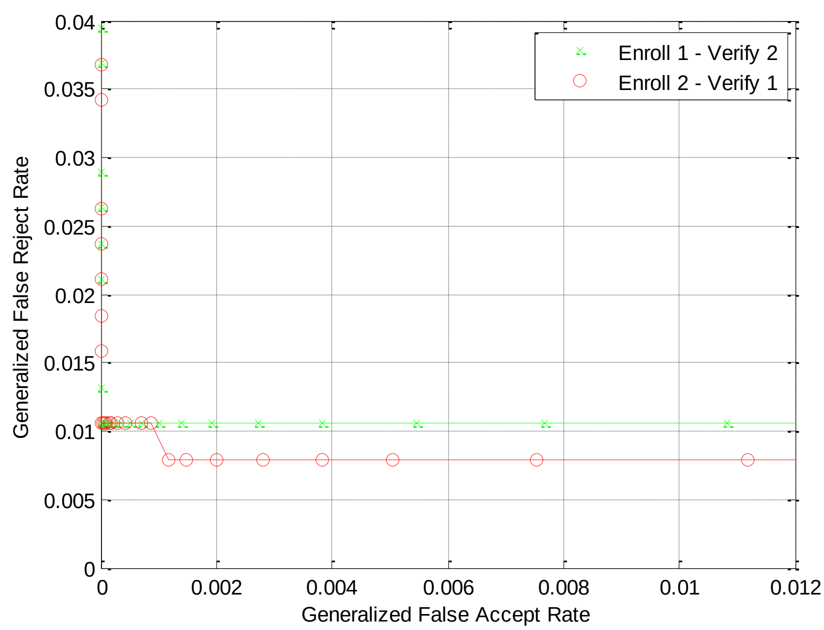


Figure 101. Enrol H Verify K – GFRR Versus GFAR

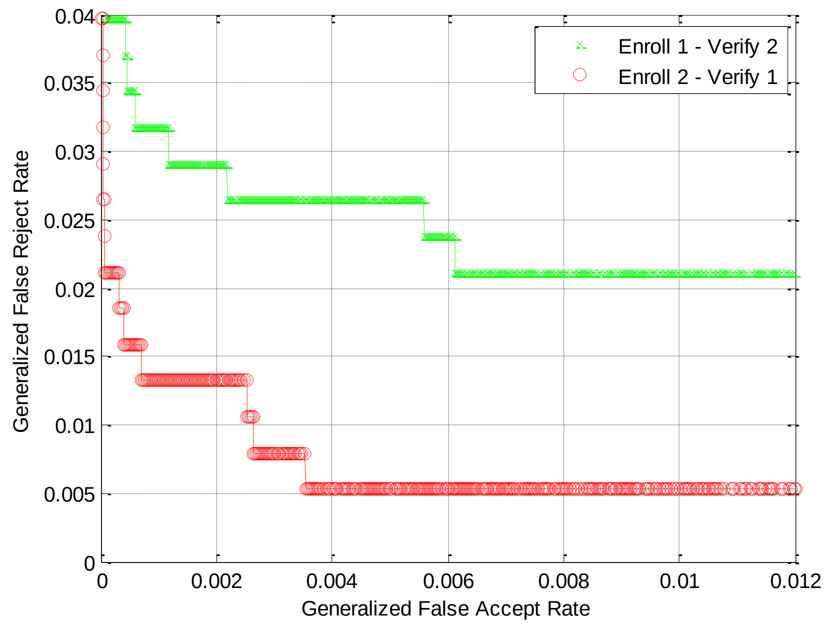


Figure 102. Enrol H Verify L – GFRR Versus GFAR

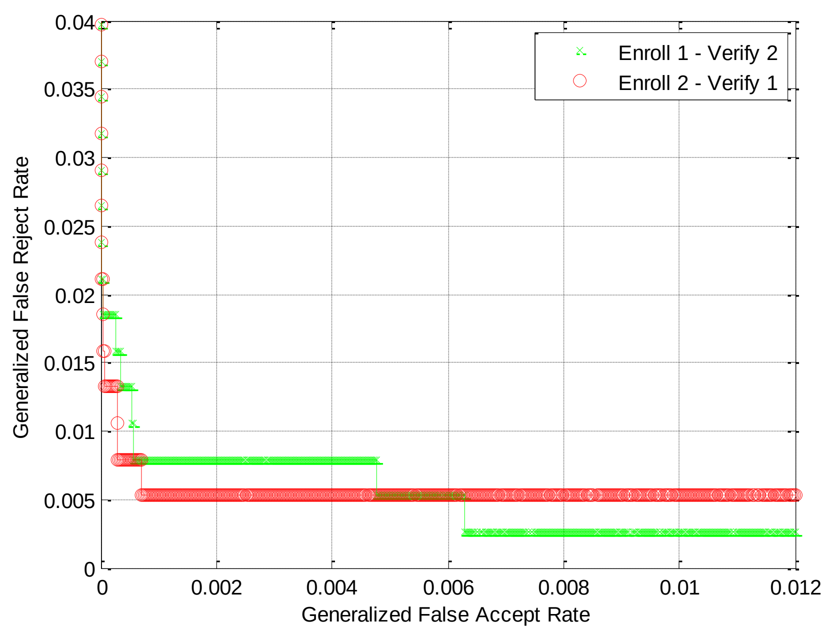


Figure 103. Enrol I Verify A – GFRR Versus GFAR

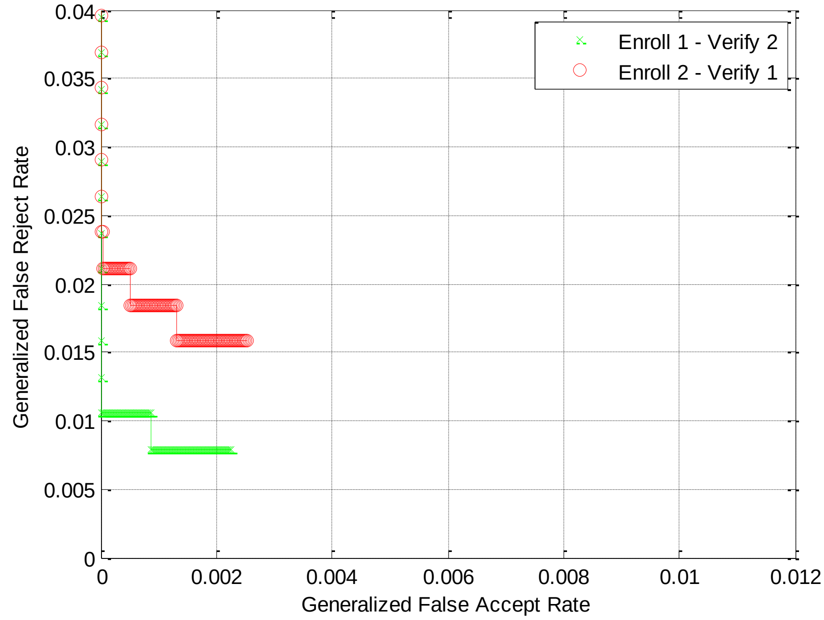


Figure 104. Enrol I Verify B – GFRR Versus GFAR

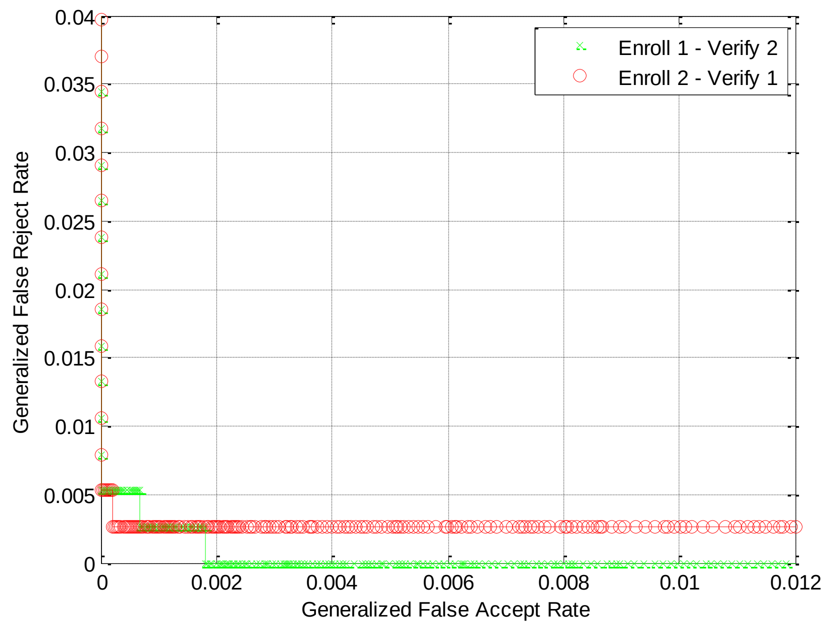


Figure 105. Enrol I Verify C – GFRR Versus GFAR

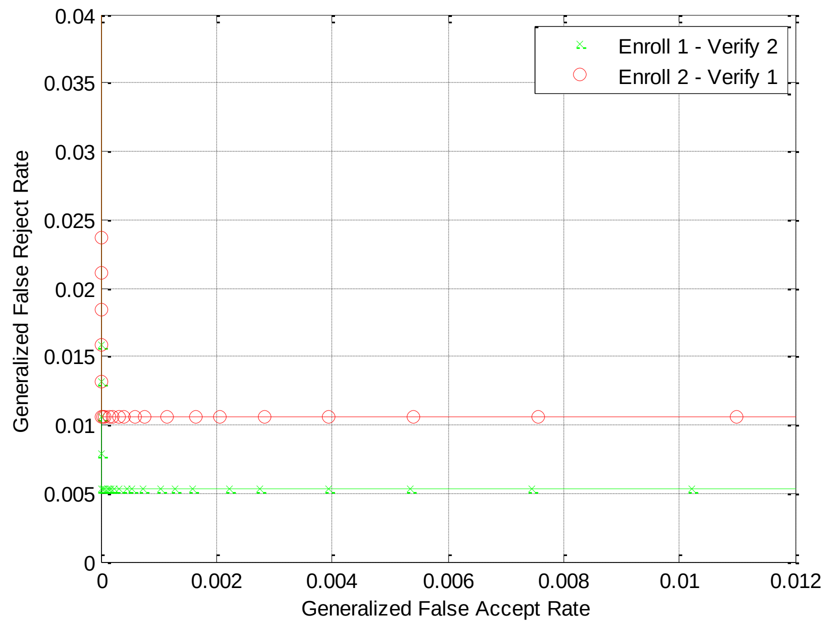


Figure 106. Enrol I Verify D – GFRR Versus GFAR

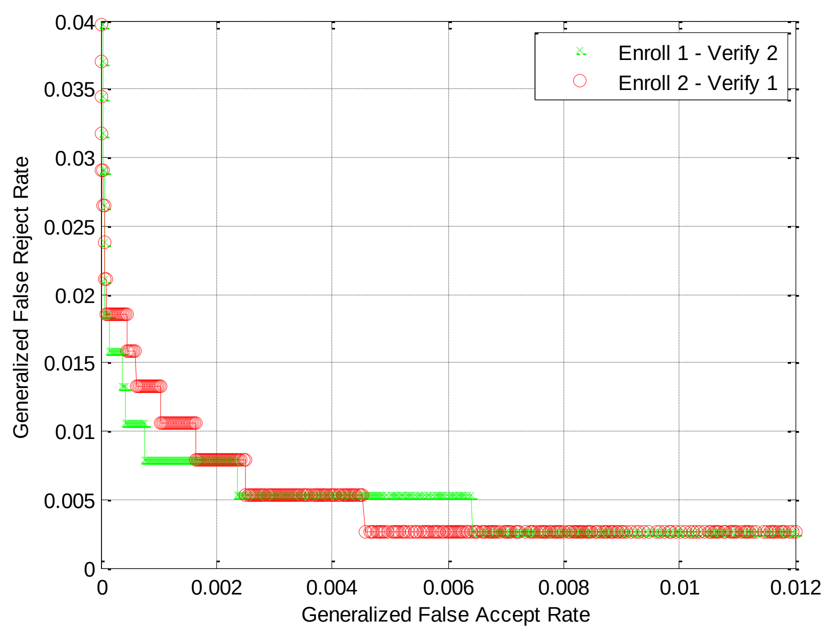


Figure 107. Enrol I Verify E – GFRR Versus GFAR

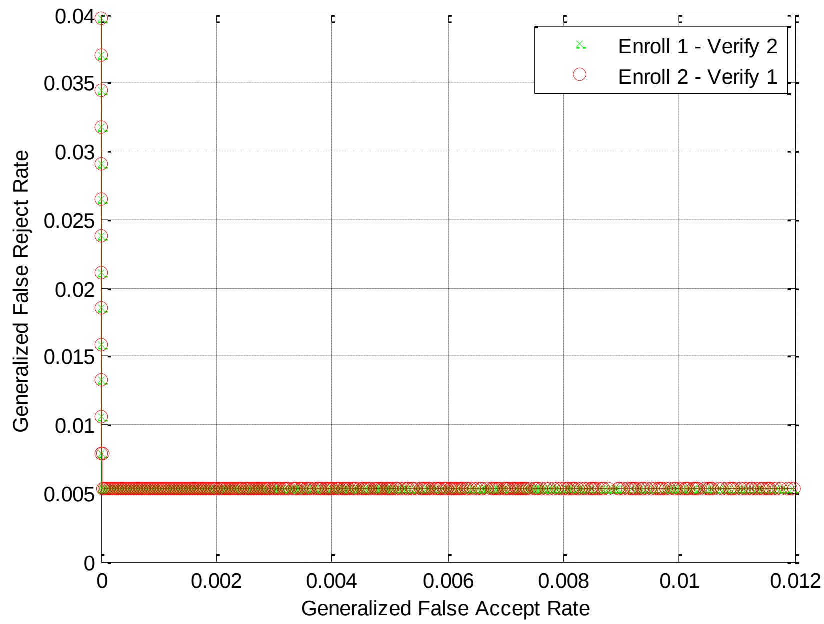


Figure 108. Enrol I Verify F – GFRR Versus GFAR

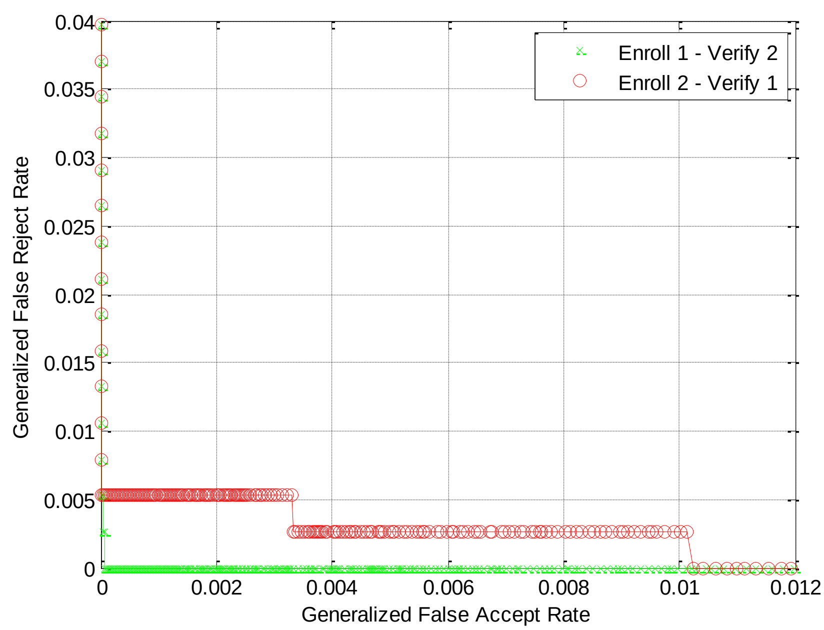


Figure 109. Enrol I Verify G – GFRR Versus GFAR

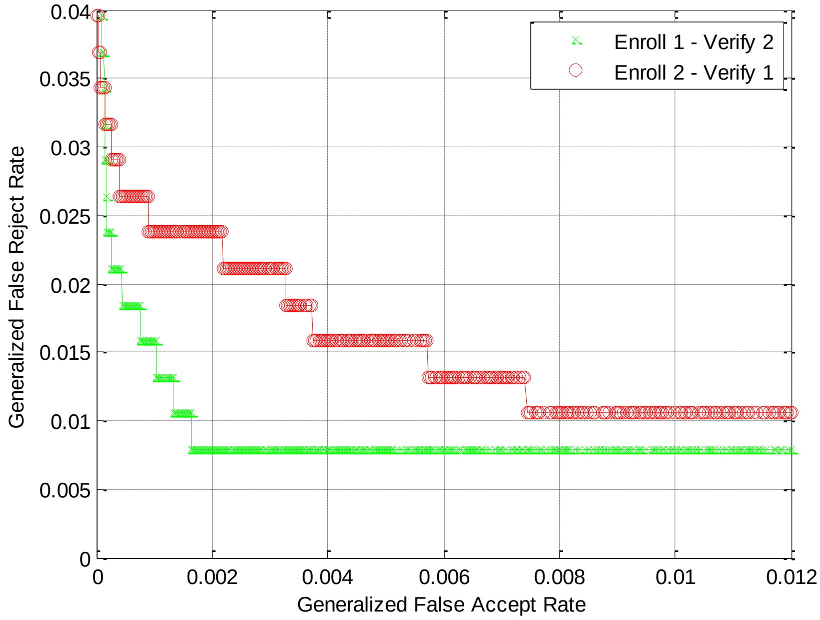


Figure 110. Enrol I Verify H – GFRR Versus GFAR

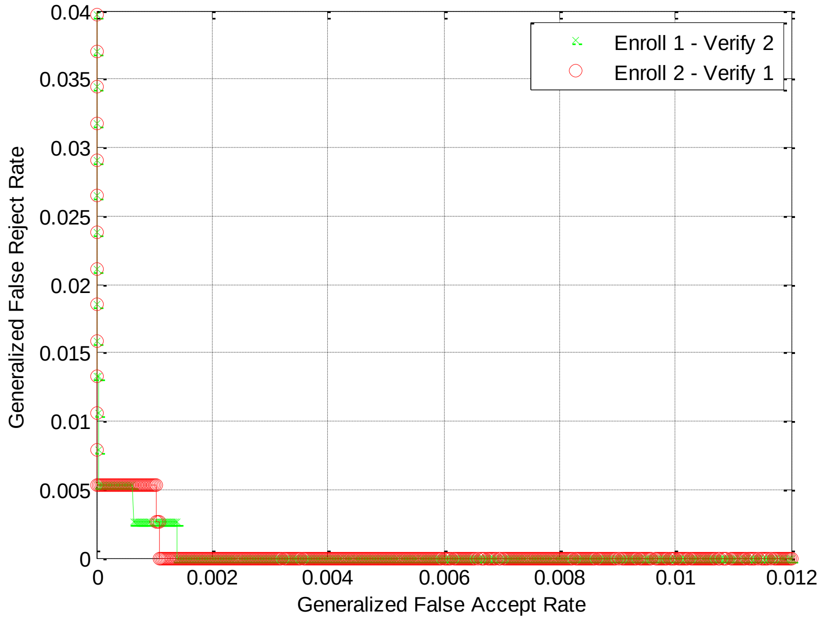


Figure 111. Enrol I Verify I – GFRR Versus GFAR

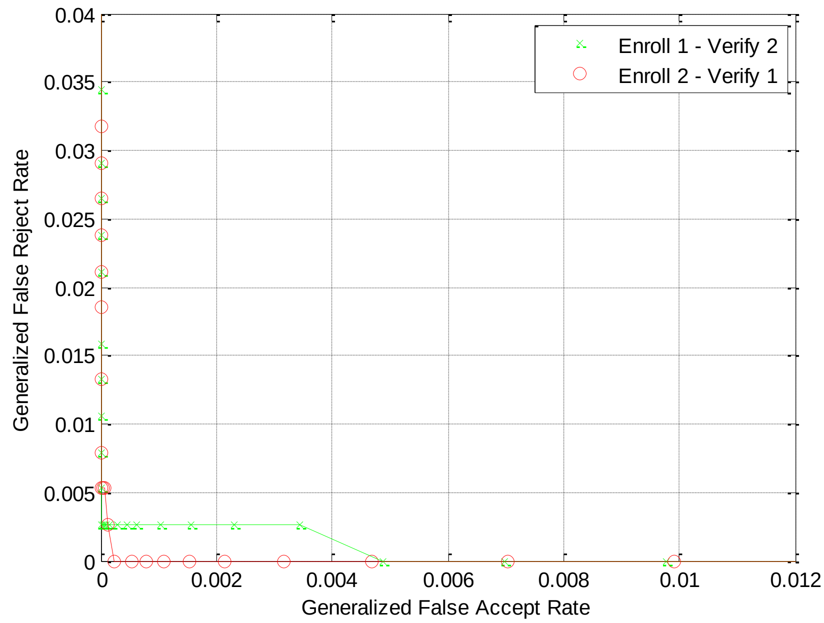


Figure 112. Enrol I Verify J – GFRR Versus GFAR

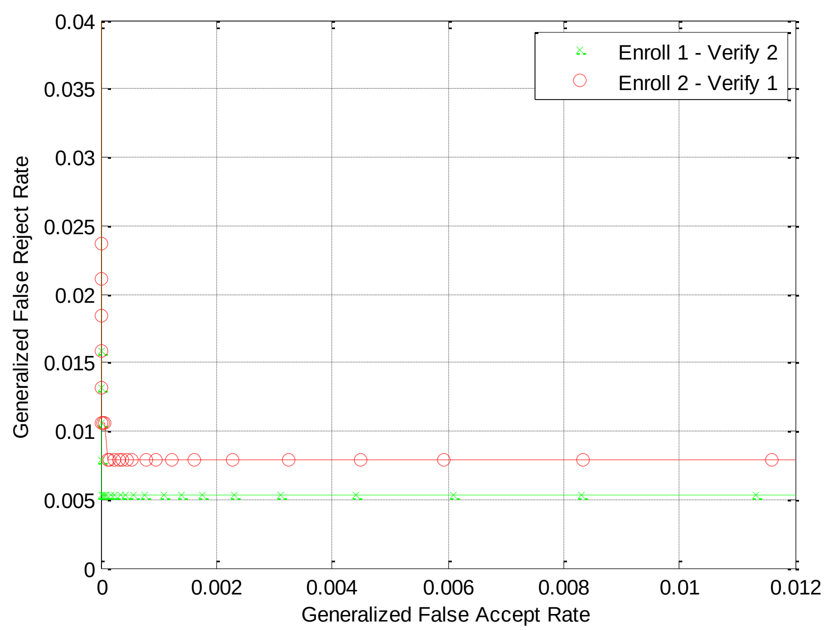


Figure 113. Enrol I Verify K – GFRR Versus GFAR

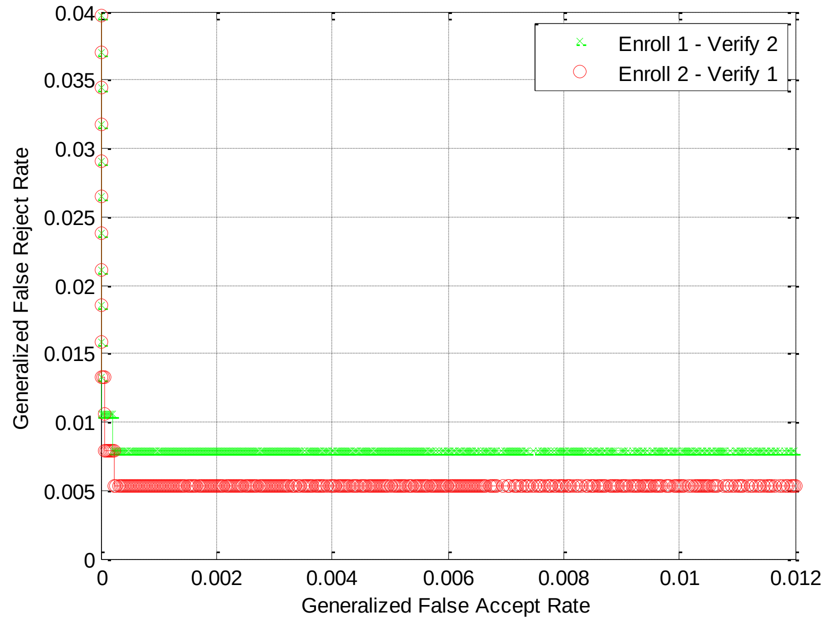


Figure 114. Enrol I Verify L – GFRR Versus GFAR

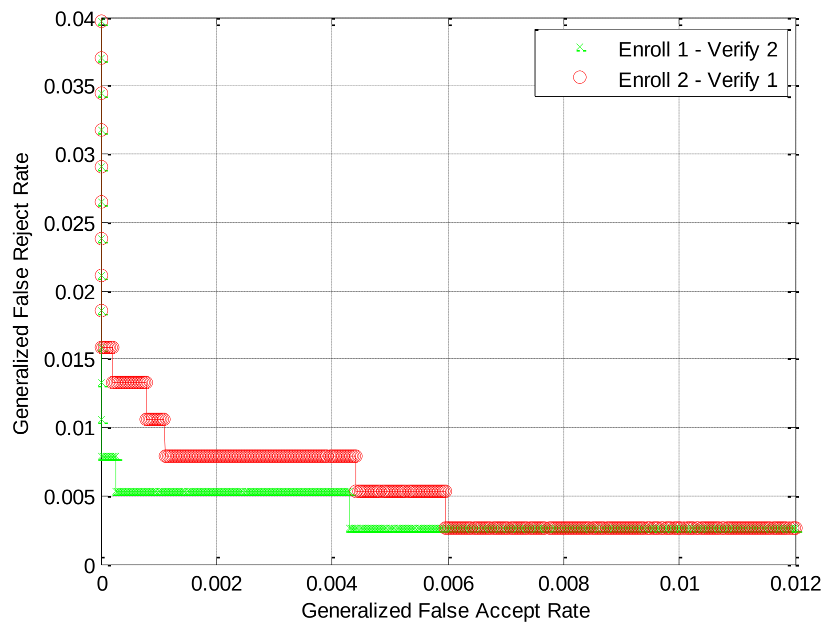


Figure 115. Enrol J Verify A – GFRR Versus GFAR

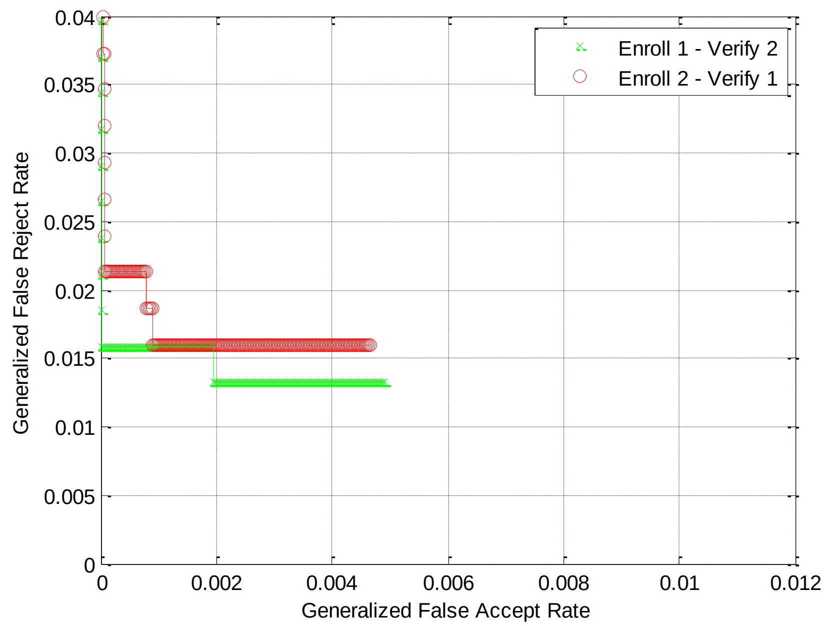


Figure 116. Enrol J Verify B – GFRR Versus GFAR

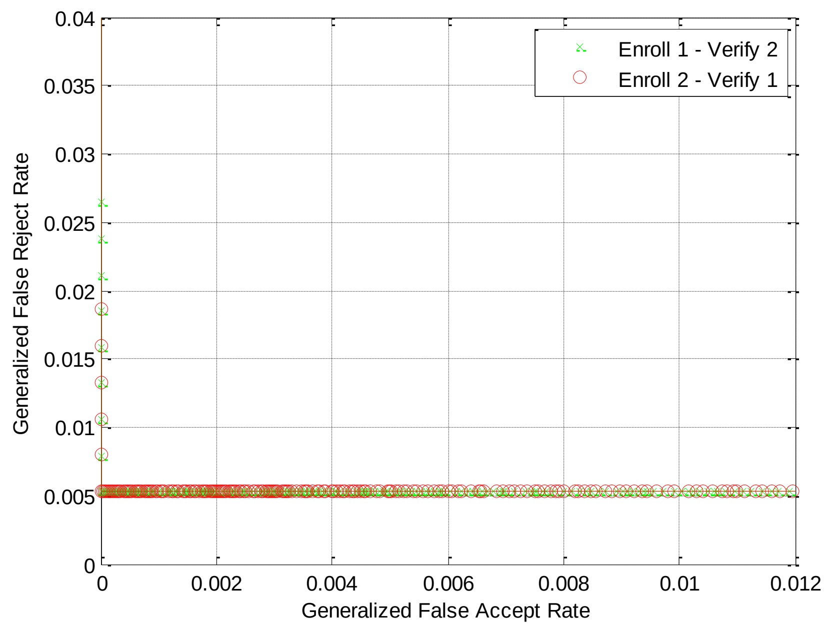


Figure 117. Enrol J Verify C – GFRR Versus GFAR

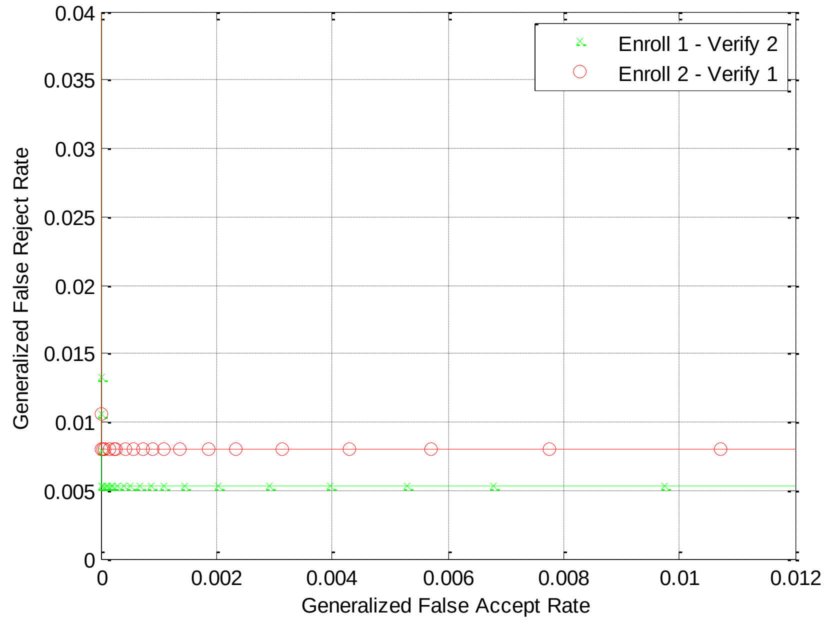


Figure 118. Enrol J Verify D – GFRR Versus GFAR

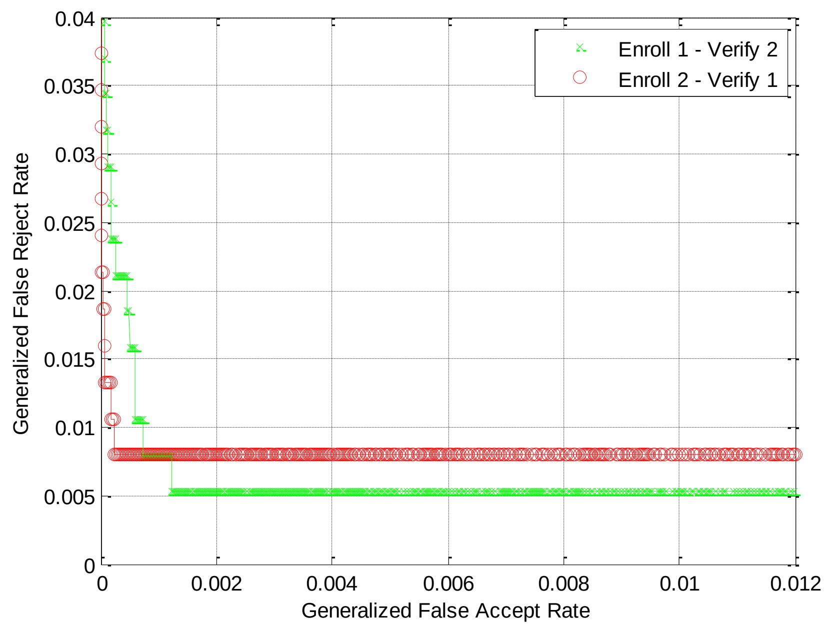


Figure 119. Enrol J Verify E – GFRR Versus GFAR

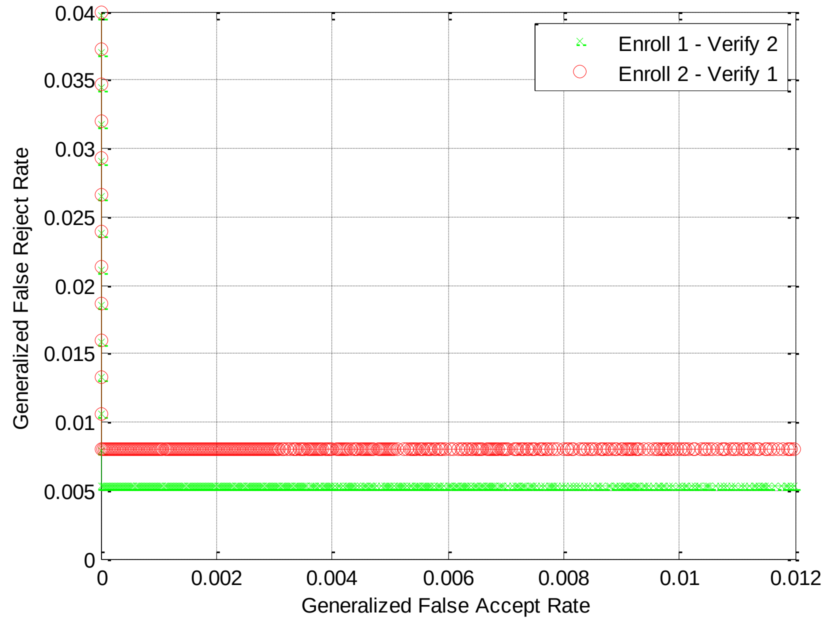


Figure 120. Enrol J Verify F – GFRR Versus GFAR

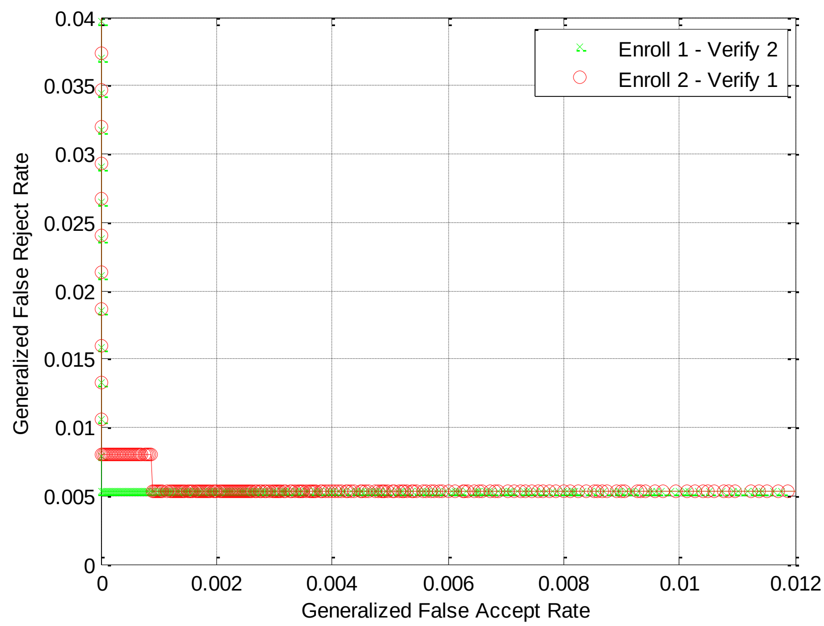


Figure 121. Enrol J Verify G – GFRR Versus GFAR

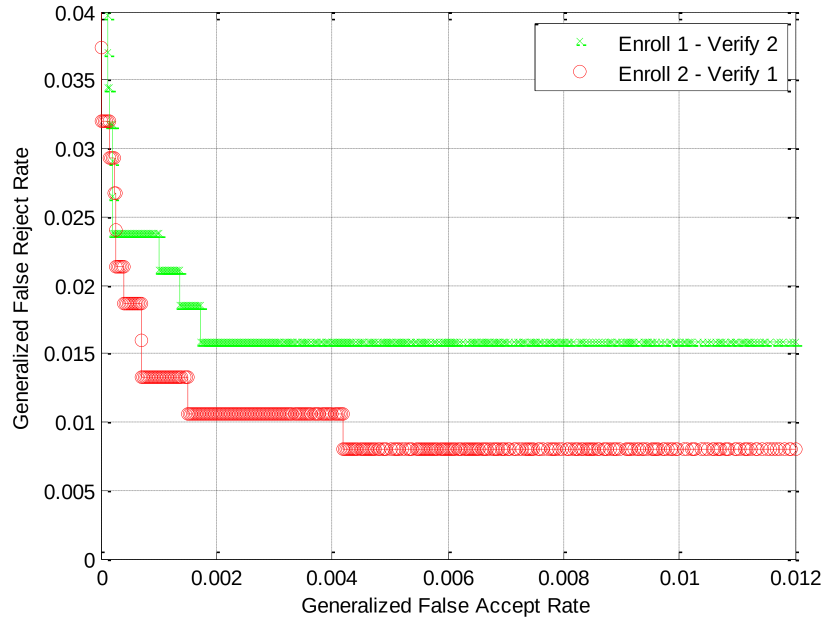


Figure 122. Enrol J Verify H – GFRR Versus GFAR

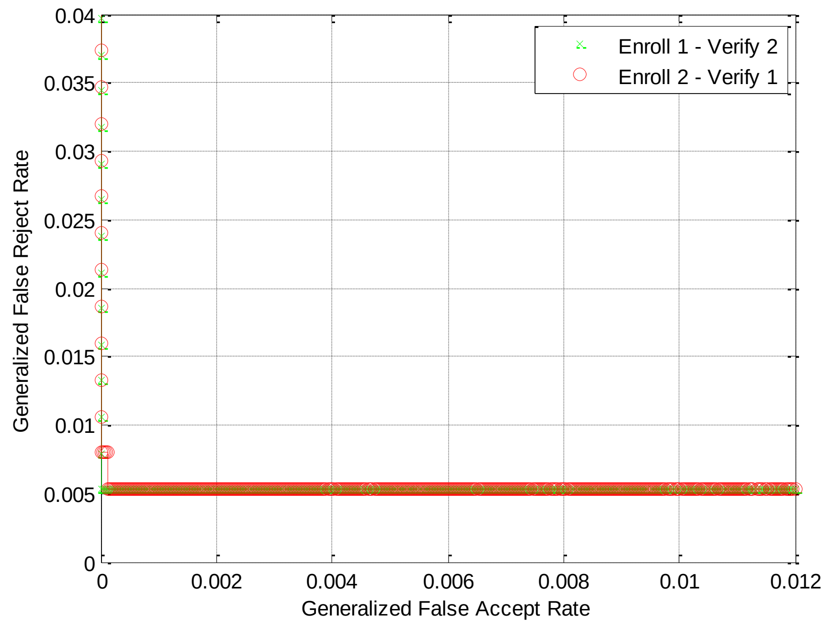


Figure 123. Enrol J Verify I – GFRR Versus GFAR

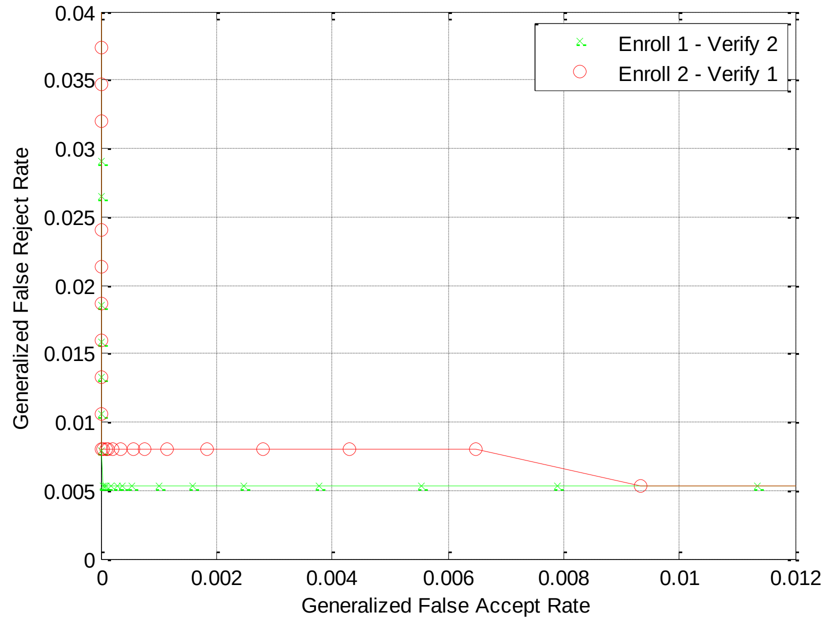


Figure 124. Enrol J Verify J – GFRR Versus GFAR

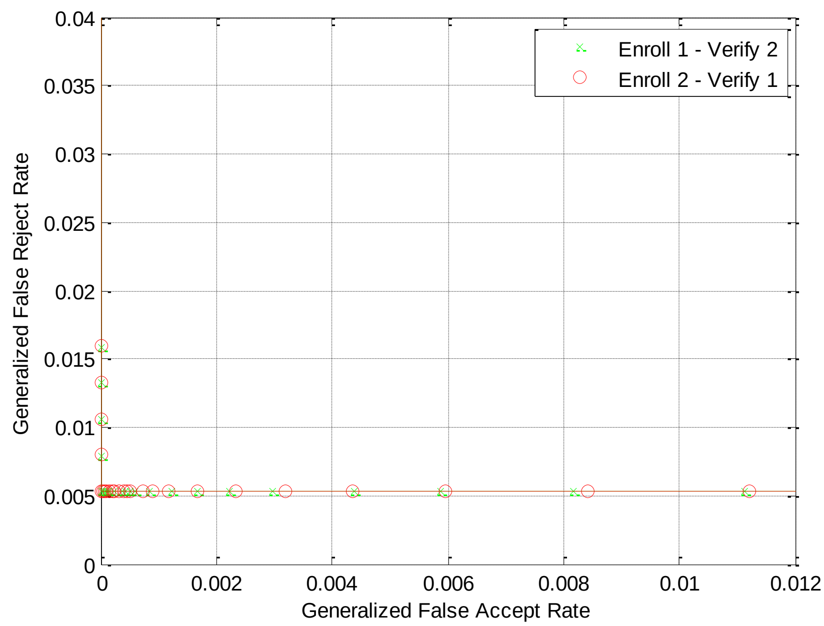


Figure 125. Enrol J Verify K – GFRR Versus GFAR

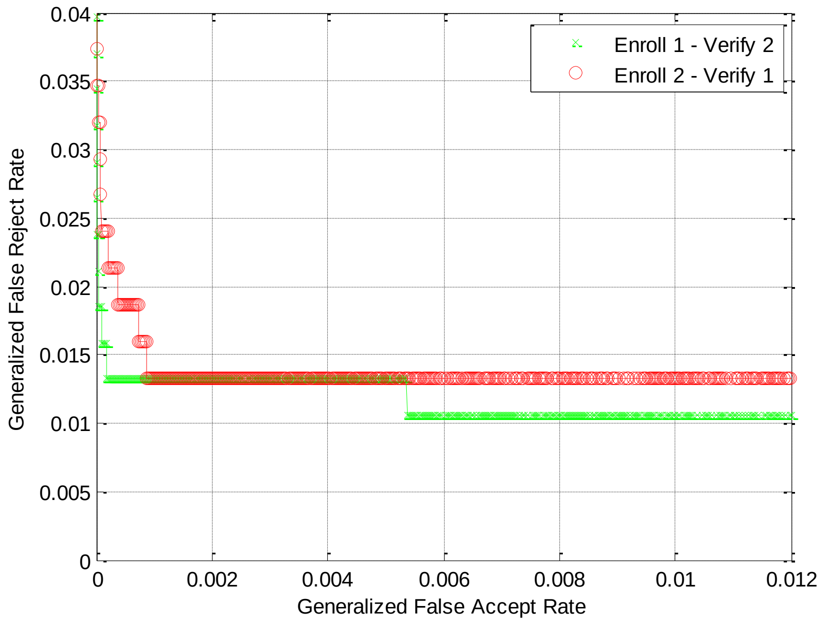


Figure 126. Enrol J Verify L – GFRR Versus GFAR

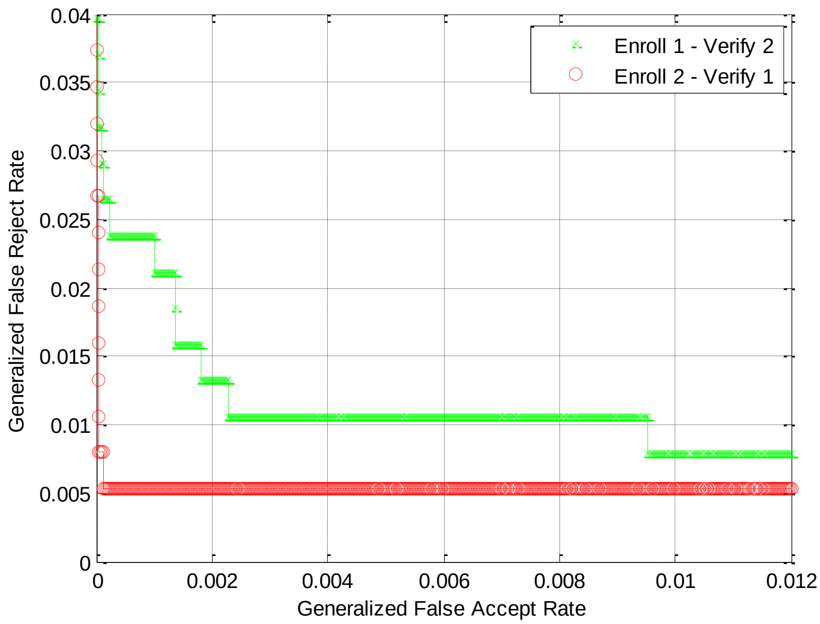


Figure 127. Enrol K Verify A – GFRR Versus GFAR

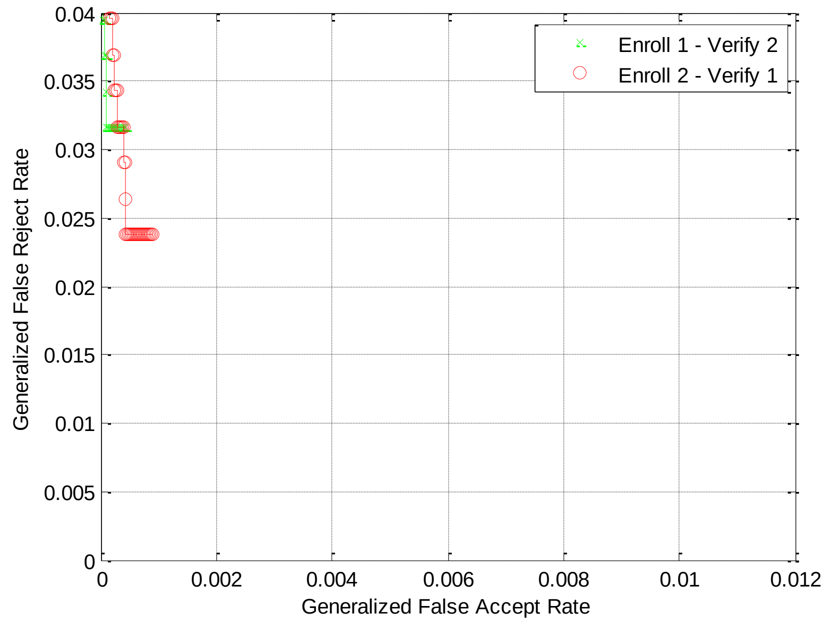


Figure 128. Enrol K Verify B – GFRR Versus GFAR

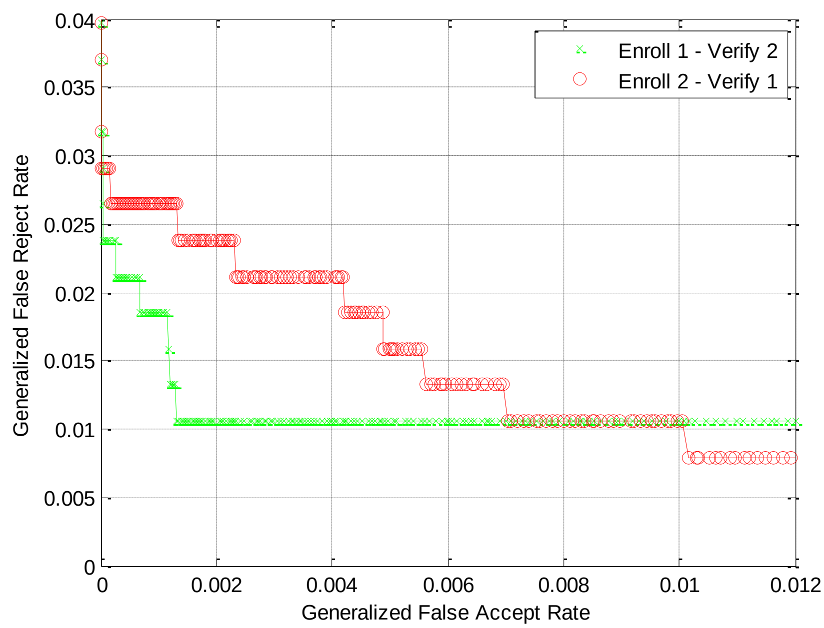


Figure 129. Enrol K Verify C – GFRR Versus GFAR

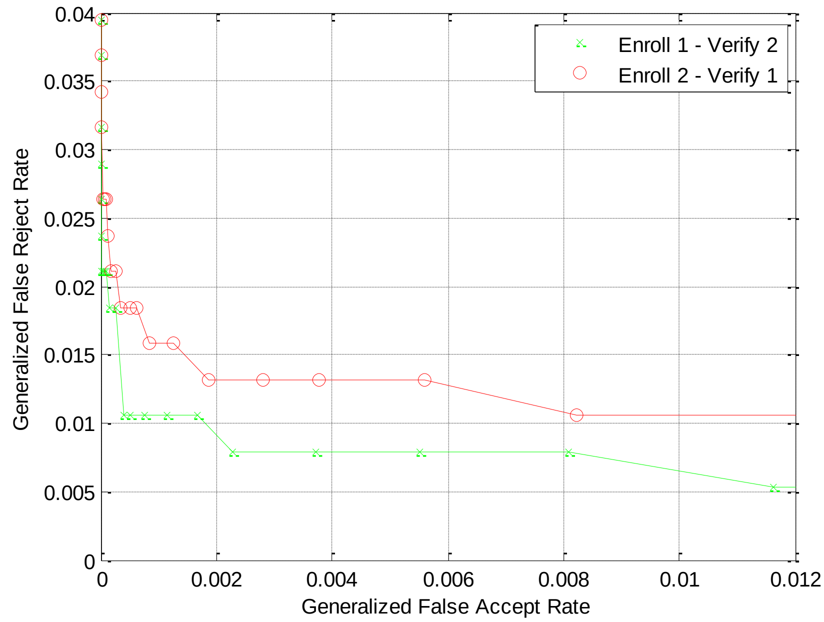


Figure 130. Enrol K Verify D – GFRR Versus GFAR

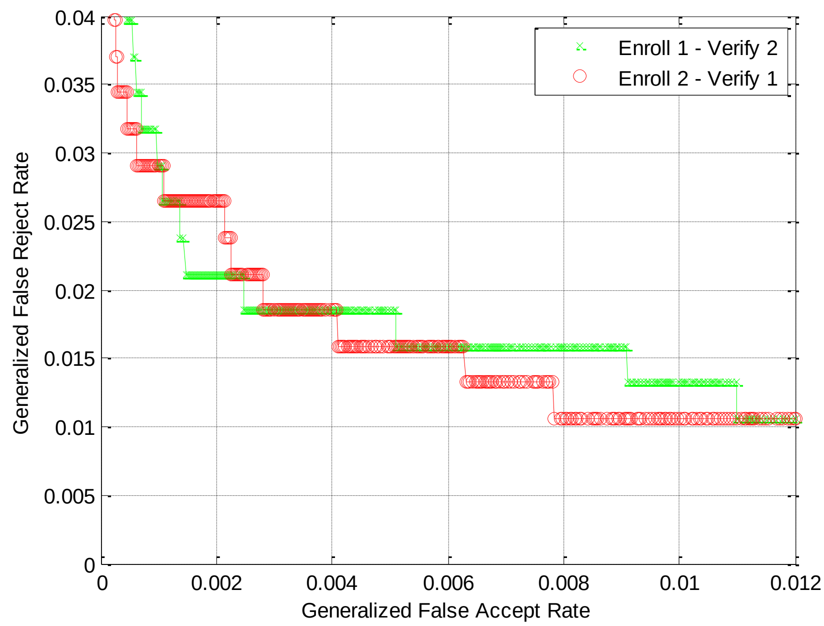


Figure 131. Enrol K Verify E – GFRR Versus GFAR

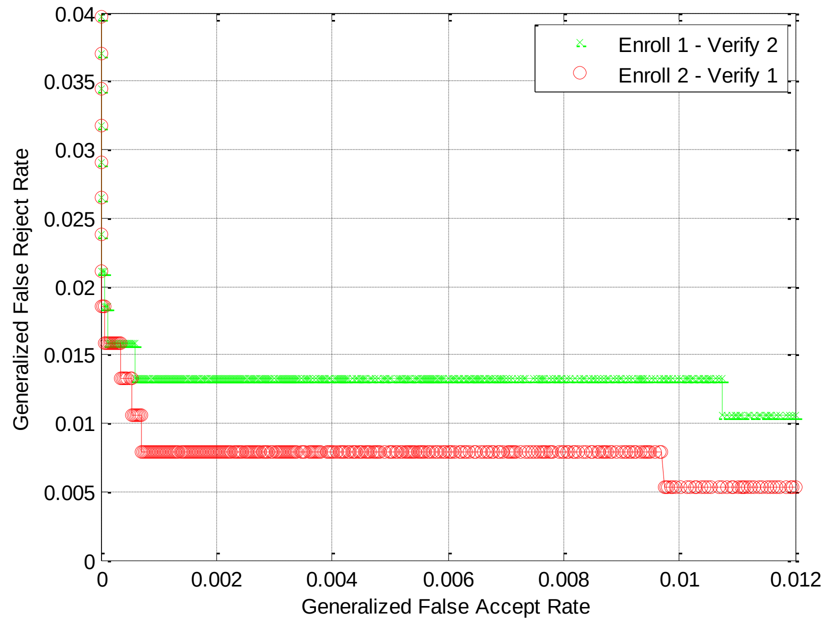


Figure 132. Enrol K Verify F – GFRR Versus GFAR

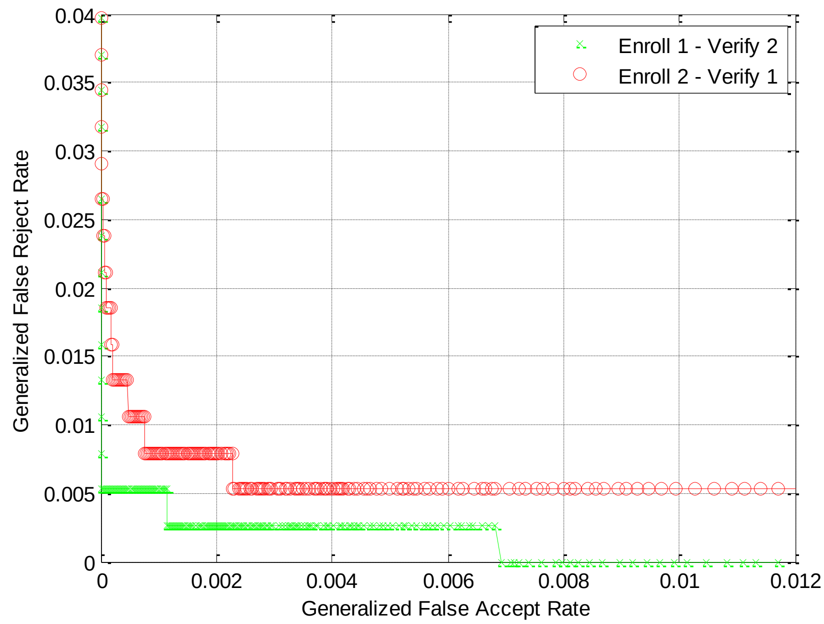


Figure 133. Enrol K Verify G – GFRR Versus GFAR

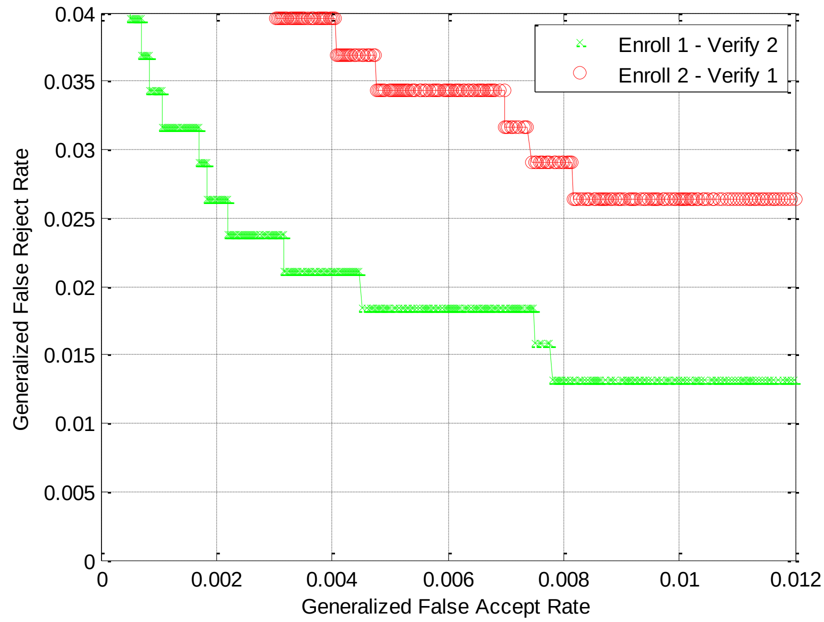


Figure 134. Enrol K Verify H – GFRR Versus GFAR

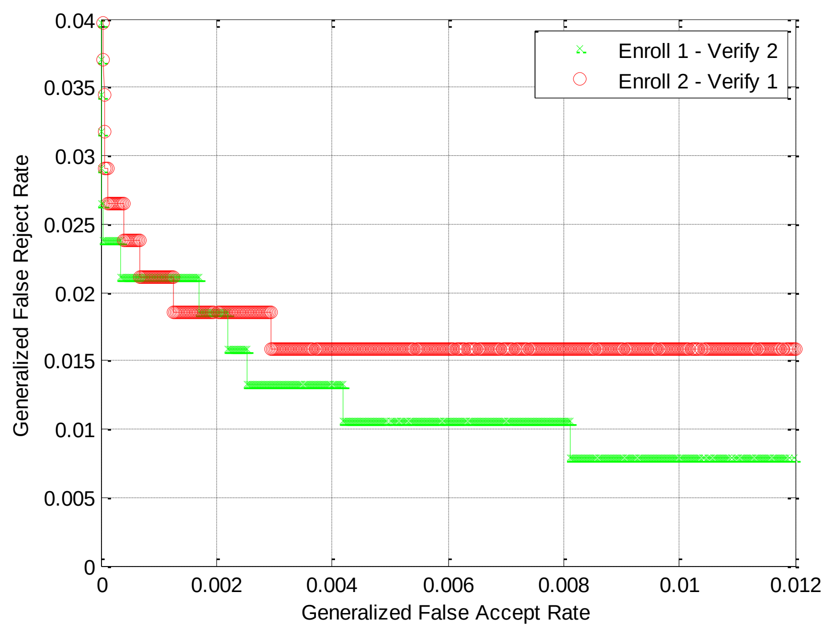


Figure 135. Enrol K Verify I – GFRR Versus GFAR

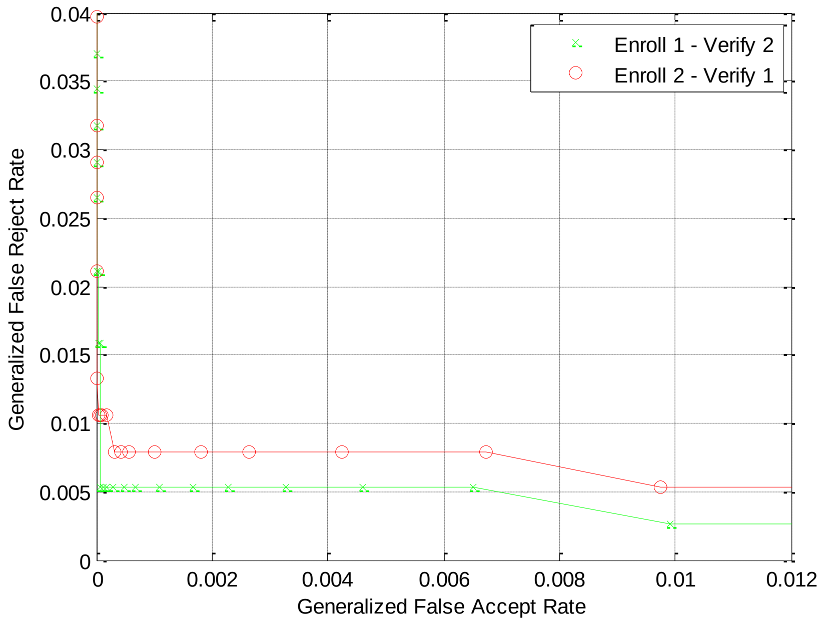


Figure 136. Enrol K Verify J – GFRR Versus GFAR

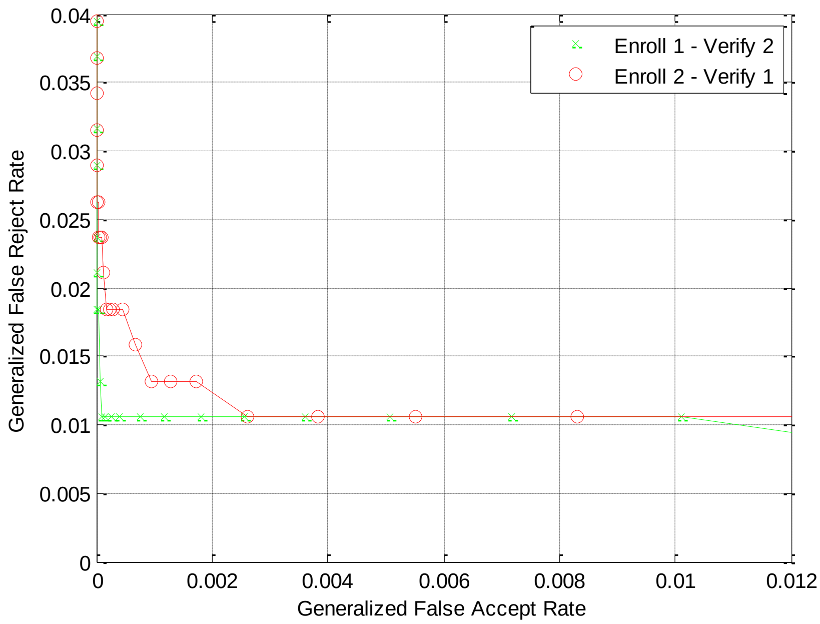


Figure 137. Enrol K Verify K – GFRR Versus GFAR

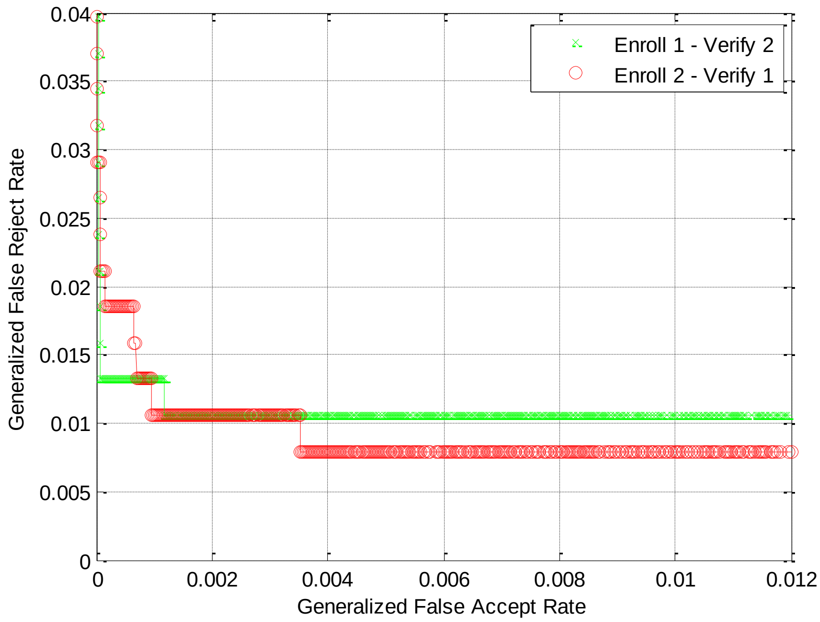


Figure 138. Enrol K Verify L – GFRR Versus GFAR

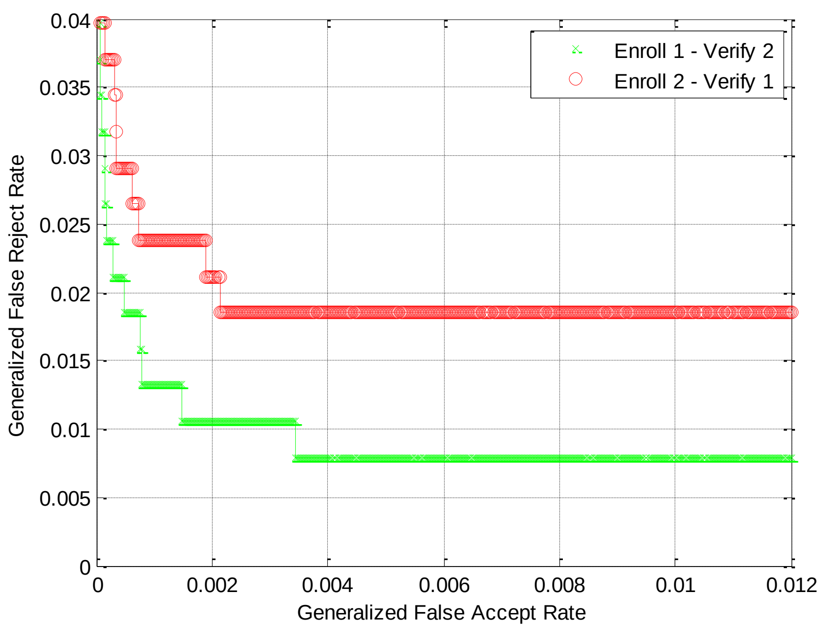


Figure 139. Enrol L Verify A – GFRR Versus GFAR

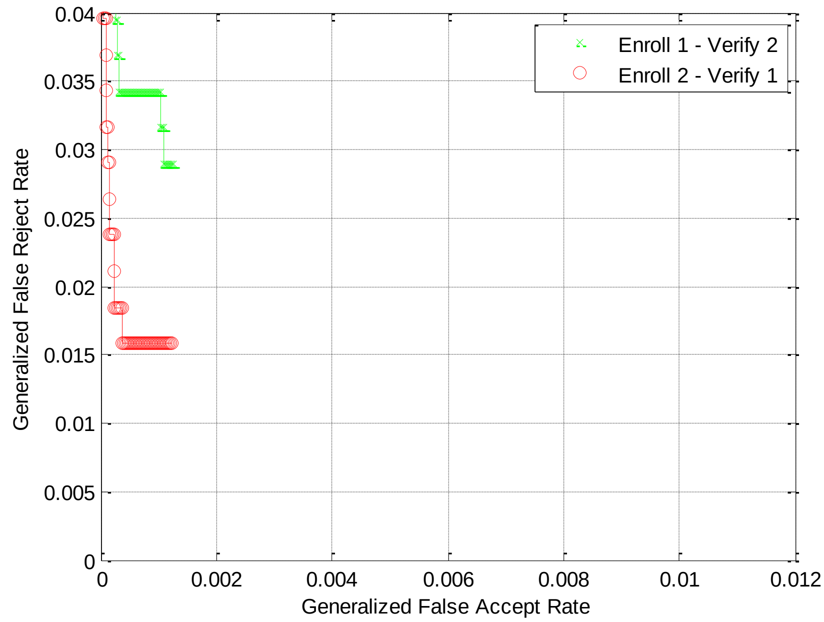


Figure 140. Enrol L Verify B – GFRR Versus GFAR

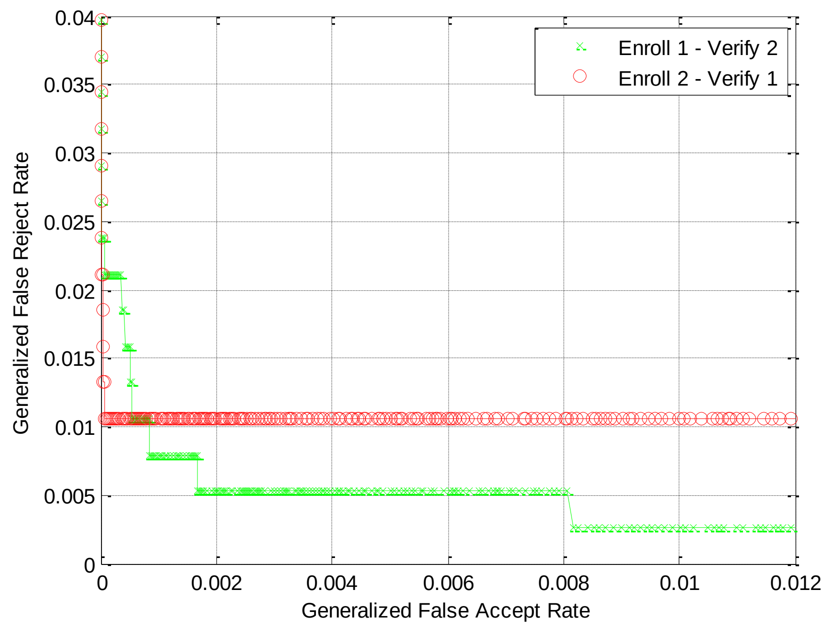


Figure 141. Enrol L Verify C – GFRR Versus GFAR

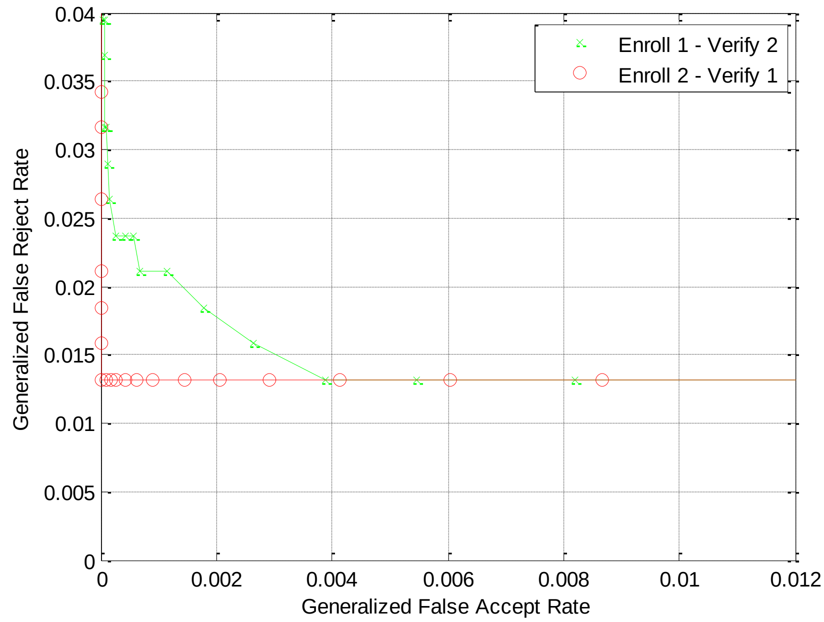


Figure 142. Enrol L Verify D – GFRR Versus GFAR

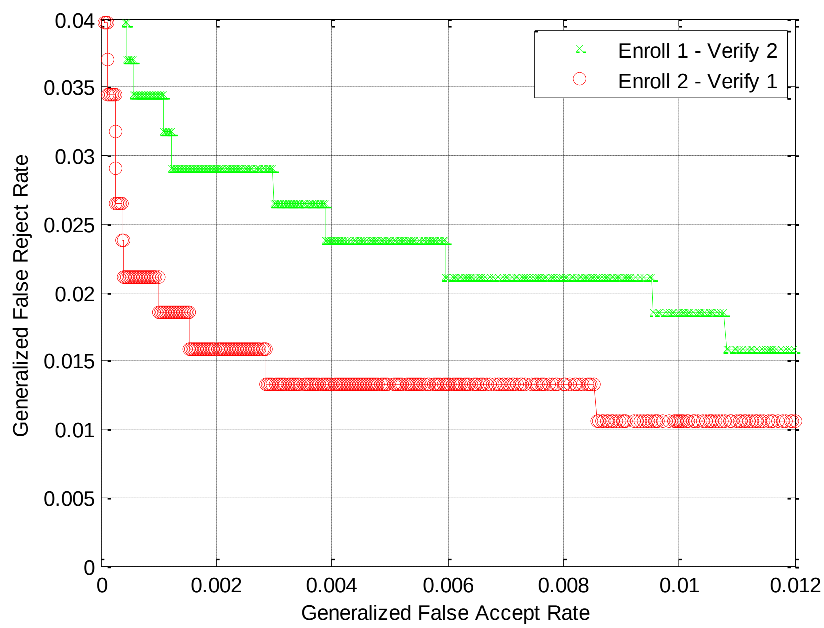


Figure 143. Enrol L Verify E – GFRR Versus GFAR

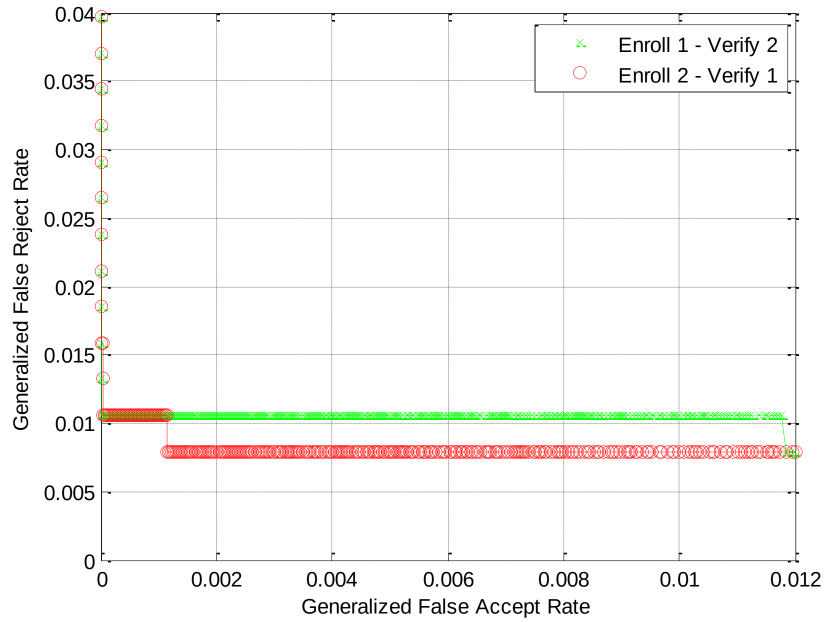


Figure 144. Enrol L Verify F – GFRR Versus GFAR

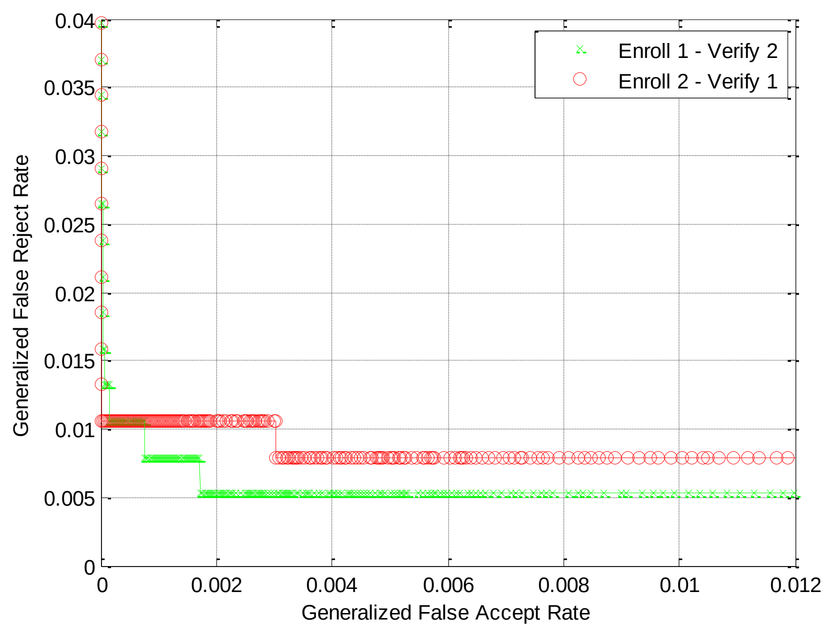


Figure 145. Enrol L Verify G – GFRR Versus GFAR

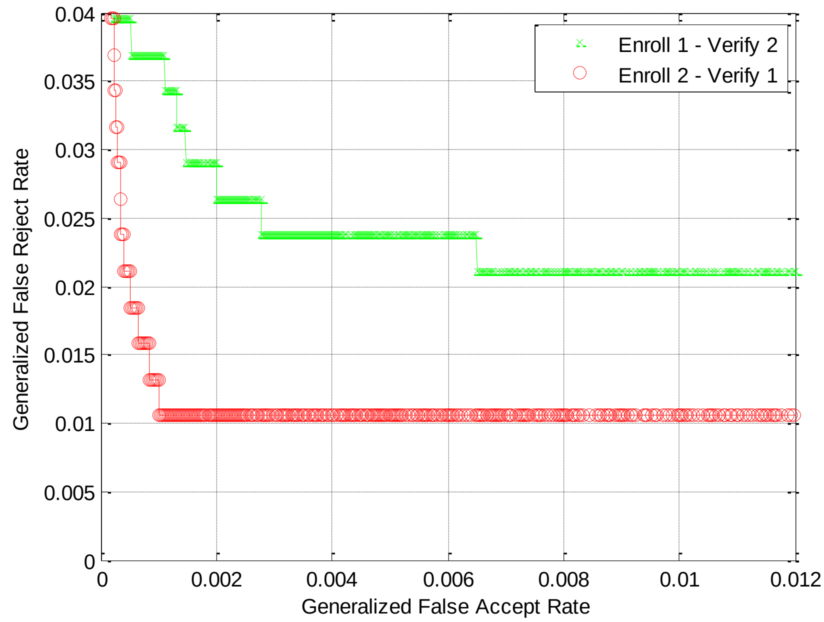


Figure 146. Enrol L Verify H – GFRR Versus GFAR

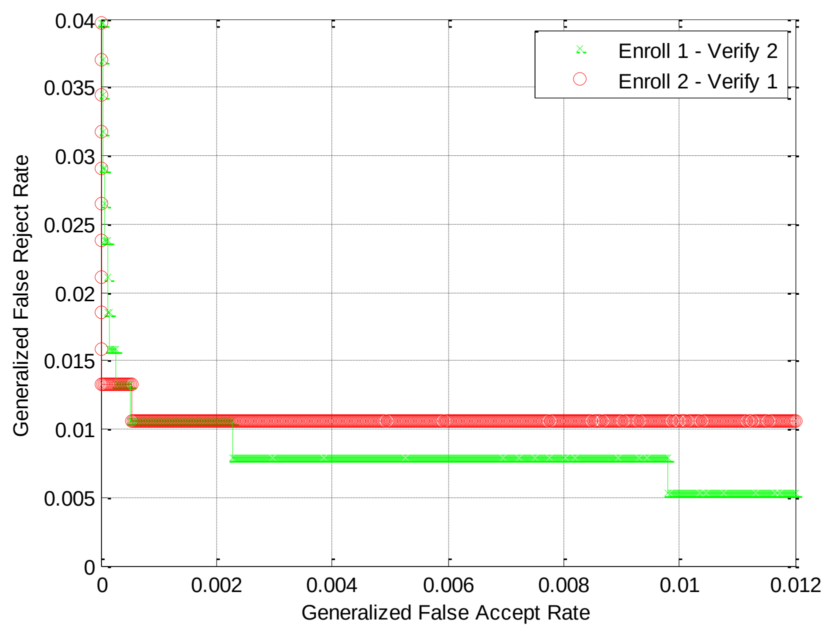


Figure 147. Enrol L Verify I – GFRR Versus GFAR

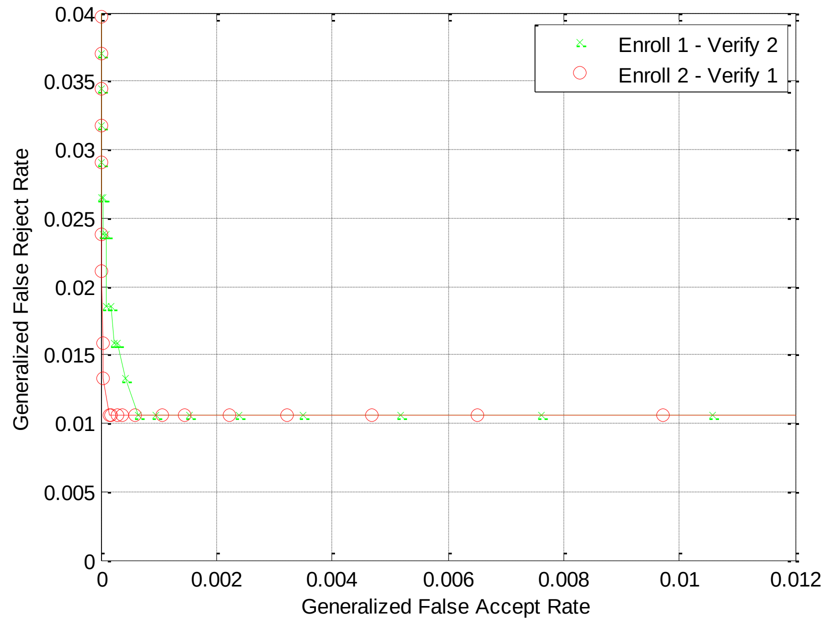


Figure 148. Enrol L Verify J – GFRR Versus GFAR

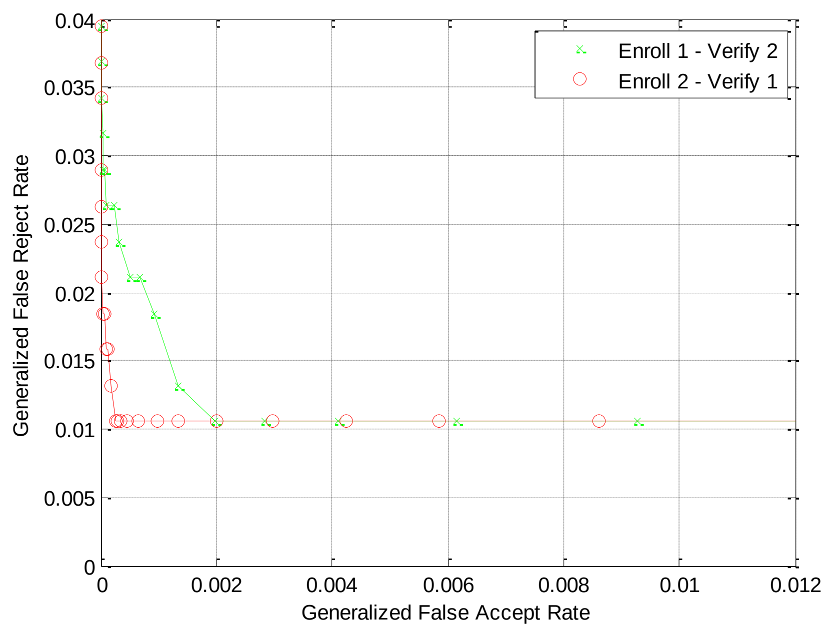


Figure 149. Enrol L Verify K – GFRR Versus GFAR

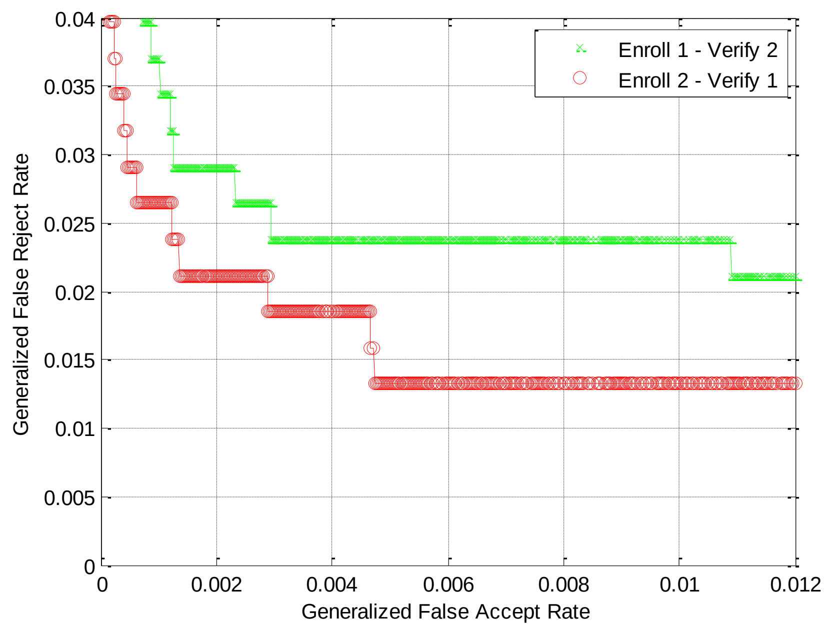


Figure 150. Enrol L Verify L – GFRR Versus GFAR