INTERNATIONAL LABOUR ORGANIZATION

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

ILO Seafarers' Identity Documents Biometric Interoperability Test (ISBIT-3) Report

Geneva, 2009

INTERNATIONAL LABOUR OFFICE GENEVA

978-92-2-123170-7 (print)
978-92-2-123171-4 (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-3**) Report

First 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). Although this report uses letters to refer to the products, ILO has published the names of the three conformant and interoperable products on its website in a list of products.

As more countries have begun investigating ILO SID deployments, more companies have become interested in having their products tested for conformance and interoperability so that they can be added to the ILO product list. The ISO/IEC 19794-2 fingerprint minutiae based data interchange format is now a published standard and there has been significant investment within the biometrics industry in supporting that standard. Together, these two facts have led to the development of sufficient interest among the biometrics community to spur a third ILO SID Biometric Interoperability Test. That test, ISBIT-3 is the subject of this document.

Executive Summary

The ILO Seafarers' Identity Documents Biometric Interoperability Test #3 (ISBIT-3) took place in Ottawa, Canada from January to June, 2006. The three products from the existing ILO list of approved products were tested, as was a replacement for one of those products that fixed a known error in the code that had been pointed out earlier by the vendor of the product. An additional six new products were also tested. Each product consisted of a sensor paired with an algorithm capable of both enrollment and verification.

The initial phase of conformance and basic interoperability revealed that all ten products were conformant to the requirements of ILO SID-0002 and could achieve interoperability with one another on a limited set of fingers. This was completed by February, 2006.

Full data collection from 184 test subjects aged from 18 to 69 took place in February and March. Each test subject visited the test laboratory twice during that period and was enrolled on every product and verified multiple times on every product during each visit. A total of 67,802 fingerprint images were collected under controlled and supervised conditions.

Data processing and analysis took place in April and May, with the final production of this report in June. A total of 161,359,702 individual matches were computed, resulting in a total of 27,066,803 two finger transactions being simulated.

Ultimately the revised version of the previously approved product achieved better interoperable performance than the original version (as expected since the nature of the error in the code was well understood) and, with the agreement of the vendor of that product, it is recommended to withdraw the previously approved product and replace it with the revised one. The vendor has given assurances that no copies of the previously approved product will remain in use for SID purposes.

Of the other six new products tested, all six achieved the target interoperable performance metric when used in conjunction with the previously approved products (including the revised version of the one mentioned above) and with each other of a mean GFRR of 1% or less at a GFAR of 1%. Specifically the mean GFRR was 0.92%. It is therefore recommended that the list of approved products be extended to include 9 products.

Contents

Index of Tables

Index of Figures

1 Introduction

1.1 Background

Fingerprint matchers based on minutiae points have been a mainstay of the biometrics industry since its infancy. In the past, different vendors typically used an internal minutiae format, commonly known as a proprietary template, to store the sets of minutiae and associated features they extracted and used for matching. This prevented interoperability, since a template created by a product from one vendor could not generally be correctly interpreted by other vendors' products and thus enrollment and verification had to use equipment supplied by the same vendor. One way to resolve this problem was to exchange the entire fingerprint image, but this was not always possible, due to storage, network bandwidth or privacy issues. Recently, work began to determine the basic elements comprising minutiae templates and formalize them in a standardized minutiae template. This was first accomplished in ANSI INCITS 378:2004 Information Technology – Finger Minutiae Format for Data Interchange, a US standard published in 2004. Later, in 2005, the International Organization for Standardization (ISO) published ISO/IEC 19794-2:2005 Biometric Data Interchange Formats - Part 2: Finger Minutiae Data. This standard is very similar to the US version, but it also supports card formats that allow the minutiae to be stored in a more compact form.

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' identity 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 defines the standard for the biometric template to be stored in the bar code and the method for enrollment and verification of the seafarers' fingerprints. The minutiae format selected was a draft of the SC 37 standard ISO 19794-2 Biometric Data Interchange Formats – Part 2: Finger Minutiae Data dated October, 2003 and the specific format selected was the normal sized finger minutiae card format ("card normal format"). The document was subsequently updated based upon the results of the ISBIT-1 and ISBIT-2 tests described below, and the revised document was approved in November, 2005.

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 enrollment 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 enrollment 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 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 single-finger live finger verification transactions, each consisting of up to 3 single finger presentations. This test is known as **I**LO **S**eafarers' Identity 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. 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). Although this report uses letters to refer to the products, ILO has published the names of the three conformant and interoperable products on its website in a list of products.

1.2 Current Test

Since the original ISBIT-1 and ISBIT-2 tests, ISO published the final version of ISO/IEC 19794-2. There was also a dialogue between the ILO and ISO/IEC JTC-1 SC 37 on the results of the interoperability tests, which will eventually lead to a revision of this standard to improve interoperability. More and more biometric vendors and end users have realized the importance of standards compliance and interoperability. Consequently, by the end of 2005, many additional vendors wished to submit products to be tested for inclusion on the ILO list of products that are compliant to the biometric requirements of Convention No. 185. Bion Biometrics was selected by the ILO as the single laboratory to manage a testing program to maintain the list of compliant products and so in November, 2005 Bion announced a testing program known as ILO Seafarers' Identity Documents Biometric Interoperability Test #3 (ISBIT-3).

The schedule required companies to submit products by January 17th, 2006 at the latest, so that the initial conformance testing phase and interaction with the vendors to modify conformant products could be completed by February 7th, 2006. Live capture of biometric data from test subjects could then occur over five and a half weeks from February 8th, 2006 until March 18th, 2006, with the offline matching using the data acquired during the live capture to begin immediately after the live capture phase was finished. In actuality, the vendors were allowed to make final modifications to the enrollment and matching software (but not the capture software) during the first week of live capture, so there were some interoperable combinations in the first week where user feedback of a match or non-match response was not representative of the performance that would be achieved with the final product. This did not affect the offline calculations that were used to derive the final interoperability performance.

The goal 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 enroll and verify product). Although ISBIT-1 had featured one enrollment session and three verification sessions, only the final verification session, in which the test staff were permitted to provide guidance to the test subjects if they were having difficulty in finger placement, was used in computing the final interoperable performance numbers. Therefore ISBIT-3 dispensed with the other sessions and only used verification sessions with assistance from the test staff. To increase the amount of data acquired with only two visits to the test centre by the test subjects, ISBIT-3 enrolled the test subjects during both visits. The enrollment session from the first visit and the verification session from the second visit (known as E1V2) were used together, as were the enrollment 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 enrollment and verification separated by two to three weeks, even if in one case enrollment was before verification and in another, enrollment was after verification. The user feedback of match or non-match for a given finger placement was always based on the enrollment for the first session which meant a slight difference of user feedback from the offline results used to calculate the final performance for the E2V1 session but this was considered an acceptable tradeoff to double the amount of usable data. The differences in performance among these two sets of data are discussed in Sections 4.2 and 4.4.

The test included the three existing approved products, as already published on the ILO list, as well as six new products from six new biometric vendors. The performance criterion that had to be met in order for a new product to be added to the interoperable products list was very simple:

The mean FRR at a FAR 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 FRR at a FAR 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. Although there have recently been developments in the ISO/IEC JTC-1 SC 37 related to a draft standard on performance based interoperability testing (ISO/IEC 19795-4) that suggested that a maximum criterion may be more suitable than a mean criterion for large groups of interoperable products, ILO had already established this criterion and it was deemed inappropriate to change it at this point.

One issue did arise, however, that forced the nature of the test to be changed slightly from that originally envisioned. The vendor of one of the three previously approved products found that they had a bug in their quality assessment code. This bug was also present in the version of the software that was approved in the ISBIT-1 and ISBIT-2 tests. Although it did not have a significant effect in those tests, which used data collected in a warm and relatively humid environment on board a ship sailing in the tropics, the bug was expected to cause problems for dry fingers. Since approved products are not permitted to be changed in any way after they have been tested, there was no simple way to fix this obvious bug that had been found by the vendor. The solution was to submit a new version of the vendor's code as a completely new product. Since the capture device was the same, there were still only 9 products used during capture and the feedback provided to the users was based on the vendor's original approved product. In the offline portion of the test, both the vendor's new code and the original code were tested and an initial set of test results were produced using the three qualified products plus the new software from one of the three vendors. An agreement was made with the vendor that if the new product did prove to have better results (as they expected it to) then the previously approved product would be withdrawn and the new product would replace it. The vendor of the product agreed to ensure that no copies of the original product or of SIDs created with it would remain in circulation if it was replaced by the new product. These three products would then be considered to form the interoperable list at the start of this test.

This situation arose because a vendor of an existing approved product found a bug in their product and desired to fix it. Since a bug free product is definitely desirable for implementers of SID systems, ILO was supportive of the vendor fixing this bug; and since it has only been a short time since the initial list of approved products was produced, it was still possible for the vendor to ensure that all copies of the original product in circulation could be withdrawn. The revised product with the bug corrected still had to be tested before it could be accepted as a replacement for the original product. In the future, as larger numbers of ILO systems become deployed and there are significant numbers of cards in circulation, it will not be possible to withdraw an existing approved product as soon as a replacement is tested and approved. It is likely that vendors of approved products may still wish to update them as their algorithms or hardware changes, but in most cases a successful test result for an updated product will result in it being added to the approved products list without the withdrawal of the previously approved product.

Although the current test featured supervised live enrollment and verification similar to that of ISBIT-1 and offline matching similar to ISBIT-2, it was structured specifically to allow for the correct types of data to be collected during the live capture portion so that the offline portion could maximize the number of two finger transactions generated. The specific methodology used to achieve this will be explained in subsequent sections, but Table 1 summarizes the impact of these changes and compares the relative size of ISBIT-1, ISBIT-2 and ISBIT-3.

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

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

[1] One product had two variants tested, but this was treated as a separate test between the three original approved products and the new variant of one of the original approved products. The final performance interoperability matrix is therefore a 9 by 9 matrix, even though matches were computed for 10 products.
[2] 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.

2 Test Methodology

2.1 General Test Conditions

2.1.1 Environment

Online capture occurred in a "normal office environment," under indirect fluorescent lighting, during the months of February and March in Ottawa, Canada. To compensate for the dry winter air from outside, a humidifier was used in the lab.

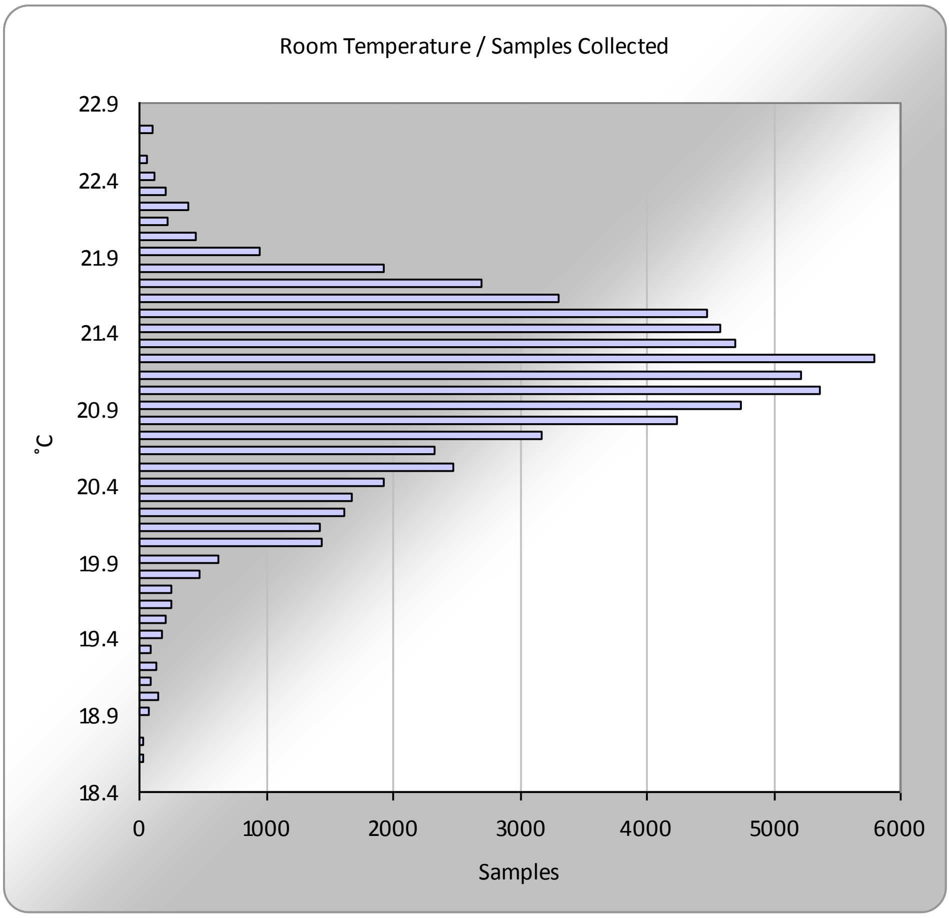
The test team made reasonable efforts to follow vendor guidance on finger placement and the use of their biometric products.

During online data acquisition, room temperature and relative humidity were sampled at regular intervals using the Extech Instruments Temperature / Humidity Datalogger 42270 / 42275. 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 21.0 ˚C, while the minimum was 18.4 ˚C, and the maximum was 22.9 ˚C.

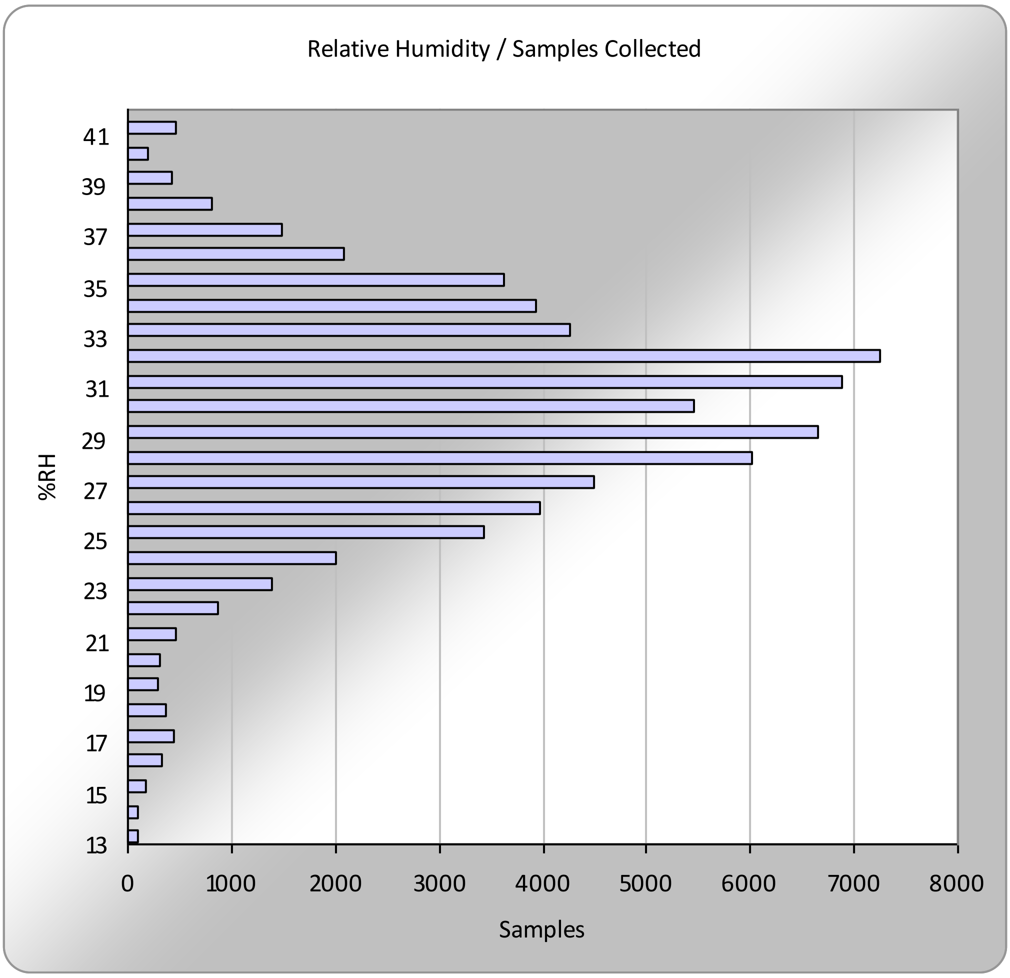
Figure 1. Samples Collected by Room Temperature



Humidity

The mean humidity for all samples collected was 30% RH, while the minimum was 13% RH, and the maximum was 42% RH.

Figure 2. Samples Collected by Relative Humidity



Note: The relative humidity was typically quite low on Monday mornings since the humidifier would exhaust the supply of water by the end of the weekend (when no subject visits were scheduled).

2.1.2 Sensor Maintenance

At the beginning of each day, each biometric sensor was wiped with an alcohol swab in a side-to-side motion. Throughout the day, sensors were wiped with a dry cloth periodically, or when an individual sensor was noticed to have residue on the scanning area.

2.1.3 Order Effects

The order in which the biometric products are used can potentially affect performance due to the reasons listed below. Therefore, the order in which products are presented to each test subject should ideally provide the same opportunity to all products to be first, second, third, etc.

- Feedback from one biometric product may affect user behaviour (e.g. finger pressure) on another.

- As each product is used, the user becomes increasingly 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.

In an effort to balance pragmatism and even distribution of these order effects for each biometric product over the duration of the online capture portion of the test, the test control software cyclically incremented the product sequence used for each test subject. The test crew occasionally deviated from that however, resulting in (the worst case) a product being used as the 'start product' 14% more often than it would have, had the test crew consistently followed the request of the test control software.

2.1.4 Product Solicitation and Integration

All vendors expressing interest in participating in the test were provided with a description of the test methodology, a detailed API specification, and any additional requirements for their products in advance of the test. With the cooperation of the vendors with the test lab, any identified issues related to integration and conformance (See Section 3) were resolved before the final test phase, when capturing fingerprints from the test crew began.

2.1.5 Test Team

The test team consisted of three members: an experimenter, administrator, and observer.

The experimenter was 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 observer guided each test subject through the enrollment and verification visits, explained the test purpose and procedures, and ensured that finger placements were done according to the administrator's instructions.

The administrator ensured that the test system functioned properly and entered test subjects' birth year, nationality group, job group, gender, and any other useful comments into the test control software. During online enrollment and verification the administrator verified image quality and placement results, while also providing finger placement instructions.

To represent supervised operating conditions, both the administrator and observer provided finger placement and quality guidance, and any available direct feedback from the biometric product, to the test subject based on their experience with the products to improve placement quality.

2.1.6 Sensor Malfunction

When experiencing biometric hardware/software crashes, errors, etc., the test team found the following workarounds useful:

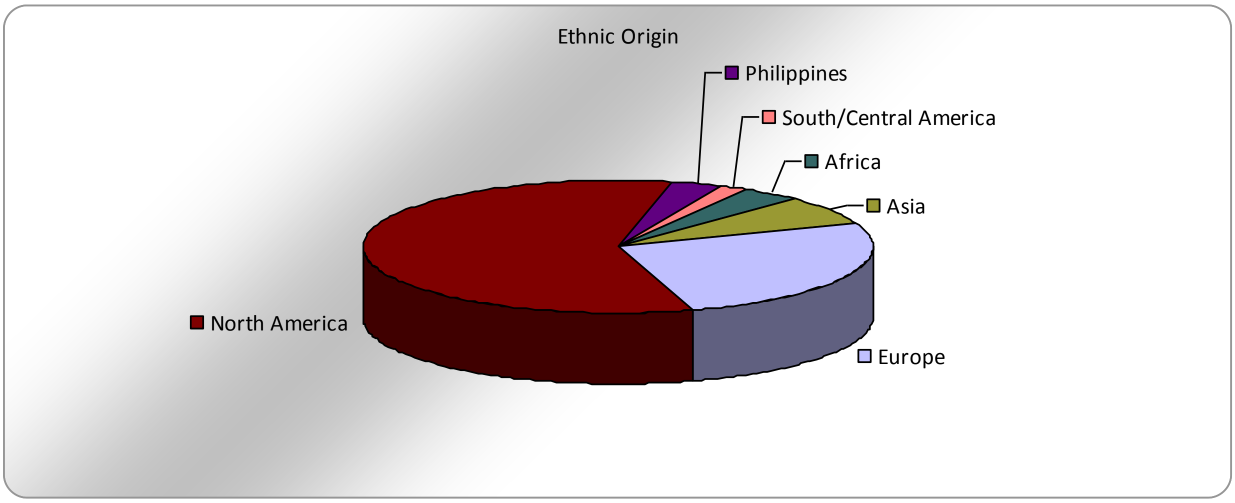
- reinitialize the sensor
- reconnect and reinitialize the sensor
- restart the test control software

- reboot the host PC

2.1.7 Test Crew

The test crew was made up of volunteers from the Ottawa area that were willing to submit their fingerprints and 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.

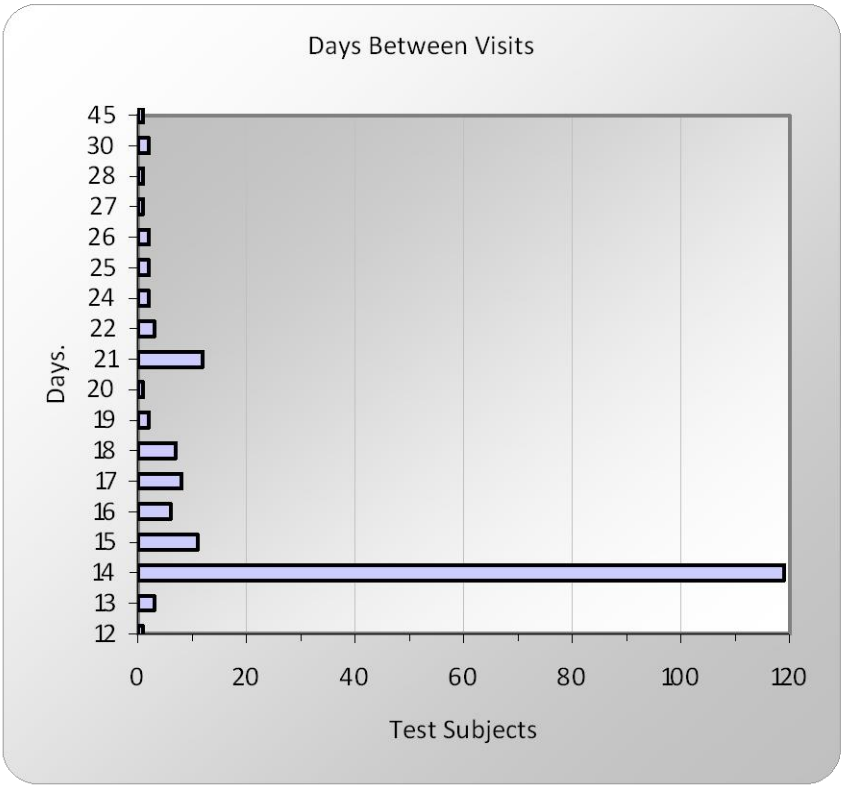
Figure 3. Ethnic Origin of Test Crew



The test crew was asked about the degree (None, Light, or Heavy) to which their work or occupation involved manual or chemical exposure. 12% of the test crew indicated heavy manual or chemical exposure, 26% indicated light exposure, while the remaining 63% indicated no exposure.

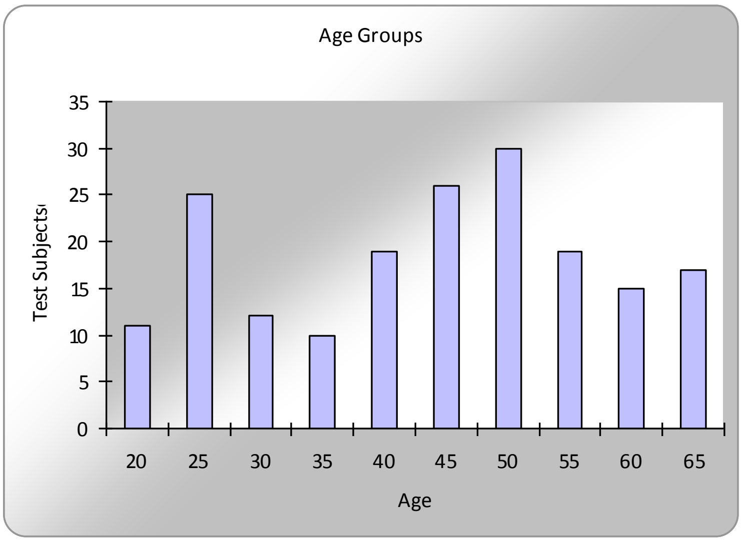
Each test subject made two visits to the test lab for the online component of the performance and interoperability test phase. Reasonable effort was made to schedule the visits at least two weeks apart. The actual mean duration between visits was 15.9 days, with a minimum of 12 days and a maximum of 45 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. There were seven test subjects aged 70 or over, but these were removed from the data as being non-representative of working seafarers, and none of the information contained in this report includes data from those test subjects. The histogram below shows the age distribution of the test subjects.

Figure 5. Age Distribution of Test Crew



Each test subject received instructions on proper finger placement via verbal instruction and physical example prior to commencing placements during their visit. Test subjects were instructed when to place a finger, and when to remove it. The administrator considered 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 removed his/her finger before being instructed to do so, the administrator would cancel and "retake" the presentation.

Throughout the visit, suggestions were offered periodically to improve the on-screen image quality on subsequent placements. 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, lotion was sometimes applied along with the occasional use of a moist cloth. When the finger was too moist, the test subject was asked to wipe the finger on a dry cloth or their clothing.

2.1.8 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 enrollment and verification
- Offline genuine comparisons
- Offline impostor comparisons
- Fingerprint image and template storage, access, and security
- Data analysis and reporting

2.2 Performance and Interoperability

The objective of the Performance and Interoperability phase of ISBIT-3 was to determine both native (enroll and verify using the same product) and non-native (enroll 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 enrollment on one product and verification on another.

2.2.1 Enrollment

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 in during that session. During enrollment, a test subject made two enrollment attempts, each consisting of three single finger presentations, to enroll 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 enroll 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 enroll on that biometric product for that enrollment visit.

The test subject was not able to participate in native genuine comparisons on that product during subsequent online verifications, although the test subject still participated in impostor comparisons and non-native genuine comparisons on that product.

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

2.2.2 Online Verification

After each visit's enrollment session, each 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 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 control software determined the unique finger positions enrolled by that test subject during their first visit for all biometric products (usually two for the right and left index), cycled the order of products used for match attempts, and randomized the match attempts for each product.

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 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 presentations per product per visit. Using this process virtually guaranteed that images would be available for exhaustive native and non-native genuine comparisons offline.

Note that the manufacturers of the biometric products have 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 for this test. 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 match presentation of a test subject's finger was matched against every BIR with 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/template with templates of 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 enrollment BIR. If a BIR 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 enrollment BIR, a transaction was not simulated for that combination. If the fingers corresponding to those in the enrollment BIR 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-3 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, one product never achieved conformance to the test API in one minor way. Under certain circumstances, it returned incorrectly formatted Windows Bitmap files, which could then not be processed for enrollment or matching. A workaround was placed in 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 enroll. Thus, this minor deviation from conformance to the API specification was resolved and the product was able to proceed to the next phase of conformance testing.

The second, more critical, phase of conformance testing involved verifying that products supported the requirements of ILO SID-0002, "The standard for the biometric template required by the Convention". This specified, among other things, the format of the minutiae based biometric interchange record (BIR) to be produced during a two finger enrollment and the specific means, when it was appropriate, of recording in this BIR 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 BIR 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 BIRs 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 Enrollment

Several enrollment 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 enroll for fingerprints of insufficient quality
- successfully enrolled two fingers if two fingers of sufficient quality are available
- successfully enrolled one finger (in the event no other finger is available)

[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.

- produced BIRs that conform to the data format specified in ILO SID-0002 Annex B

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 BIRs 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 BIRs, a basic interoperability test was performed to ensure that this product could successfully work with conformant BIRs 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 standard where flexibility is allowed. One of the products may be flexible enough 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 enrollment and for verification on each product. The enrollment images were used to produce conformant BIRs 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 BIRs produced by every product. In general, if the verification component of a product's software could read and produce match scores using one BIR from a particular product then it would produce reasonable match scores for all of the BIRs produced by that product (when matching against images acquired on its own sensor from the same test subject).

Some products, however, were unable to read BIRs from certain other products. Whether this was a problem with the product that produced the BIR or with the product that attempted to read and perform verification using the BIR was difficult to determine. If a particular product had problems reading BIRs from multiple other products or produced BIRs that had difficulty being read by multiple other products, then it was likely that product was at fault. In some cases, however, there was no clear evidence which product was responsible for the interoperability failure.

The solution was to send each vendor the set of verification images for their own product and all of the BIRs produced by every product. The vendors were then able to determine for themselves why they might have difficulty reading certain BIRs. This resulted in several iterative rounds of feedback, managed through the testing laboratory, that allowed all of the vendors to modify their products so that they were both conformant and interoperable on a basic level with all of the other products.

The exact list of conformance tests performed in the first two phases of conformance testing and the results obtained by all 10 products (including both the original and revised version of the approved product that contained a quality analysis bug) are shown in Table 2 on the following page. It is apparent that they all passed all of the conformance tests, except for Product C, which had the issue described above related to invalid Bitmap files. Since this was a test harness issue and not an ILO SID-0002 conformance issue, all ten products were declared conformant and allowed to proceed to the full performance based interoperability test.

Table 2. Conformance Testing Results

Product	Reference Notes (See Below)	A0	A1	B	C	D	E	F	G	H	I
ISBIT API Tests											
CaptureInit											
	1	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
CaptureEnd											
	2	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
Capture for Enrol											
	3	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	4	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	5	Passed	Passed	Passed	Failed	Passed	Passed	Passed	Passed	Passed	Passed
Capture for Verify											
	6	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	7	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	8	Passed	Passed	Passed	Failed	Passed	Passed	Passed	Passed	Passed	Passed
Enrol											
	9	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	10	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	11	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	12	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	13	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	14	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	15	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	16	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
VerifyProcess											
	17	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
VerifyMatch											
	18	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	19	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	20	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	21	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	22	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	23	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	24	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	25	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	26	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	27	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	28	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
BIR Conformance Verification											
	29	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	30	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	31	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	32	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	33	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	34	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	35	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed	Passed
	36	Passed	Passed	Passed	Passed	Passed	Passed	Passed	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 Enrollment

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	Enrollment returns zero (0) if 6 valid images provided (I,I,I,I,I,I)
10	Enrollment returns minus one (-1) using 3 Null for Primary Finger and 3 valid for Secondary Finger (N,N,N,I,I,I)
11	Enrollment returns zero (0) using 2 Null for Primary Finger and 3 valid for Secondary Finger (I,N,N,I,I,I)
12	Enrollment returns zero (0) using 1 Null for Primary Finger and 3 valid for Secondary Finger (I,I,N,I,I,I)
13	Enrollment returns minus two (-2) using 3 valid for Primary Finger and 3 Null for Secondary Finger (I,I,I,N,N,N)
14	Enrollment returns zero (0) using 3 valid for Primary Finger and 2 Null for Secondary Finger (I,I,I,I,N,N)
15	Enrollment returns zero (0) using 3 valid for Primary Finger and 1 Null for Secondary Finger (I,I,I,I,I,N)
16	Enrollment 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 BIR from Test 9
19	Successful genuine match of secondary finger intermediate template with BIR from Test 9 with UseSecondary (See Annex A, Section A.2.8)
20	Failed genuine match of primary finger intermediate template with BIR from Test 9 with UseSecondary
21	Failed imposter match of primary finger intermediate template with BIR from Test 9
22	Failed imposter match of secondary finger intermediate template with BIR from Test 9 with UseSecondary
23	Successful genuine match of primary finger intermediate template with BIR from Test 13
24	Failed genuine match of primary finger intermediate template with BIR from Test 13 with UseSecondary
25	Failed imposter match of primary finger intermediate template with BIR from Test 13
26	Failed imposter match of primary finger intermediate template with BIR from Test 13 with UseSecondary
27	Failed genuine match of primary finger intermediate template with BIR from Test 16
28	Failed genuine match of secondary finger intermediate template with BIR from Test 16 with UseSecondary
29	BIR returned from Test 9 meets all conformance assertions
30	BIR returned from Test 10 meets all conformance assertions
31	BIR returned from Test 11 meets all conformance assertions
32	BIR returned from Test 12 meets all conformance assertions
33	BIR returned from Test 13 meets all conformance assertions
34	BIR returned from Test 14 meets all conformance assertions
35	BIR returned from Test 15 meets all conformance assertions
36	BIR 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 although this test attempted to measure the interoperable performance of a group of products that conform to “The standard for the biometric template required by Convention No. 185”, it is subject to a number of caveats. Some of these are limitations that affect all biometric tests and some of them are simply points that should be noted when considering the results. 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 composed of fingerprint sensors combined with enrollment 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. This was done when comparing product A0, the original approved product containing a quality measurement bug, to product A1, the corrected version of the product. Since the sensor and image capture software were the same for both products, the same set of images was used to test both products. In general, however, live capture is required and every member of the test crew will have to enroll 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 entice them and close supervision to ensure that they do not become fatigued and start to make mistakes when following the instructions of the biometric products. The close supervision requirement also adds to the cost by increasing the minimum number of test staff. 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 enrollment process must be used, potentially resulting in up to 30 single finger placements (3 placements of up to 10 fingers) before a failure to enroll 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.

The next caveat is that the correct treatment of failure to enroll (FTE) and failure to acquire (FTA) on a transactional basis is not well established in the literature of biometric testing. Typically, FTE and FTA are computed from single finger presentations and they can then be used to produce a generalized false reject rate (GFRR) from the false non-match rate (FNMR) or a generalized false accept rate (GFAR) from the false match rate (FMR). The formulas corresponding to a test in which each person presents a finger once for an enrollment attempt and once for a verification attempt and in which the person still makes verification attempts even if they did not enroll successfully are given below. Note that ISBIT-3 compelled each test subject to make verification attempts for every finger they had enrolled on any product, regardless of whether or not they had enrolled it on the current product, and thus a failure to enroll did not stop a test subject from making verification attempts, so these formulas would apply if the ISBIT-3 results were based on single finger presentations and not transactions.

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

$$GFRR = FTE + (1 - FTE) * FTA + (1 - FTE) * (1 - FTA) * FNMR$$

ISBIT-3 did use transactions, however, and it is necessary to consider what this means for FTE and FTA. Since a test subject was permitted to start with the right index, then the left index, then the right thumb, etc. until they worked their way through all 10 fingers in an effort to enroll at least one of them, and since three finger presentations were allowed for each finger during an enrollment attempt, a failure to enroll was rather unusual. In the previous version of ILO SID-0002, there was no way to indicate that only one finger could be enrolled, but that has been changed in the current version and so a true failure to enroll occurs only if no fingers can be enrolled. If a single finger is enrolled, verification transactions can still occur, but they will only use three placements of a single finger, rather than three placements of both enrolled fingers. Given the cold climate in Ottawa during the winter when the experiments were taking place, the fact that there were times when the humidity in the test lab was below 20% RH, and the fact that most of the products being tested used optical sensors, there were some difficulties in reading fingerprint images, and a few people did fail to enroll on a few products. There were also 2 cases during the entire experiment where an enrollment BIR was found to be non-conformant, even though the product in question produced conformant BIRs in every other recorded instance. Since there were only 2 cases, a decision was made to consider the product conformant and simply treat the non-conformant BIRs as failures to enroll. This is an excellent reminder, however, of the need to verify every enrollment BIR in a real system to ensure they are all conformant.

Failure to acquire was also affected by the environmental conditions and there were numerous times when a single finger placement was either not detected by the sensor, or the fingerprint acquisition software determined that the quality was too low and so it would not record the image. In some cases, the vendors did follow the recommended procedure in the ISBIT-3 API document and record whatever image was available when the manual cancel button was used after the 12 second time limit was exceeded. In some of those cases, the saved image could still not be processed because the enrollment or matching software determined the quality to be too low. In a very small number of cases (less than 1 match per 2 million) measured during the test, an image that was accepted for processing against a conformant BIR failed to return a match score, so although this was also a contribution towards FTA, it had a very small effect during ISBIT-3. All of the different contributions to failure to acquire mentioned above combined to produce an overall failure to acquire rate for a single finger placement. In an ISBIT-3 transaction, however, the test subject would receive a match score even if only one of their three presentations with either of their two (or one in a few cases) enrolled fingers was successfully acquired and therefore a failure to acquire was only recorded when all six (or three) presentations failed to be acquired.

In a similar manner, the FMR and FNMR were considered on a transactional basis. A false match occurred when an imposter achieved a match score above the match threshold on at least one of his or her six (or three) finger presentations during a verification transaction. Note that the choice of six or three transactions does not depend on the imposter, but on whether the genuine that they are attempting to match against was able to enroll 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 his or her six (or three) finger presentations. This could be caused by any mix of low match scores and single presentation failures to acquire. It simply did not matter when the transactional values were being considered.

Taking all of the above into consideration, it became evident that on a transactional basis, it was possible to define a transactional GFAR and GFRR using exactly the same formulas as given above for the single presentation GFAR and GFRR. The only difference was that the FTA, FTE, FNMR and FMR were all the transactional equivalents of the single presentation values. This GFRR metric is actually very useful for predicting real world performance, as it takes into account both FTE and FTA. Therefore, the metric used in ISBIT-3 to determine the list of interoperable products will be the mean GFRR at GFAR = 1% as determined from the interoperability matrix for the group of products.

The next point to consider is that the ISBIT-3 test made full use of the available live capture data by enrolling the test subjects and then attempting to verify them during both their first and second visits. During the first visit, the live feedback on match or non-match given in response to the verification presentations was based on enrollment data acquired during the same visit. During the second visit, the live feedback during verification was also based on enrollment data from the first visit. This made it more realistic, since an average of 16 days separate the first and second visit. During the exhaustive offline matching, however, the first visit enrollment records were matched against the second visit verification transactions and the second visit enrollment records were matched against the first visit verification records. This meant there was always the same amount of time between enrollment and verification, but in one half of the data enrollment 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 enrollment records might show better performance, since the test subject might be more familiar with the products during the second visit. This did have the advantage, however, that the data naturally split into two relatively independent subsets, allowing for some estimates of the confidence interval.

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-3. Most estimates of confidence require independent samples, and in the ideal case, require 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 preferably a better knowledge of the underlying population distributions. In ISBIT-3 we have obtained a mixed demographic group representing different sections of the world and different job types, but it is not clear how well that represents the overall cross-section of seafarers. We have also attempted to control environmental conditions and record what was variable, but obviously the winter conditions in Ottawa meant that most people had dry skin coming into the lab. This may well be the case for ships or ports in high or low latitudes during their respective winters, but it will certainly not be the case for tropical environments, such as that encountered during ISBIT-1. The samples are also not completely independent. Each verification transaction is composed of three or six presentations, and although the same presentations are never used in multiple transactions for the same position in the interoperability matrix, as described in Section 4.3, the same set of presentations of, for instance, the right index finger for a particular test subject, may have been used in every row for that verification product, provided that all of the enrollment products were able to enroll the right index finger. This is typical for interoperability testing, but it tends to reduce the statistical independence of the entries in the different parts of the matrices.

Despite all of these notes and caveats, the ISBIT-3 test was still a state of the art performance interoperability test and was designed to be a reasonable predictor of real world performance. Ultimately no biometric test, including ISBIT-3, can ever be a perfect predictor of real world performance, but they can come reasonably close, and this was the goal of ISBIT-3.

4.2 DET Curves

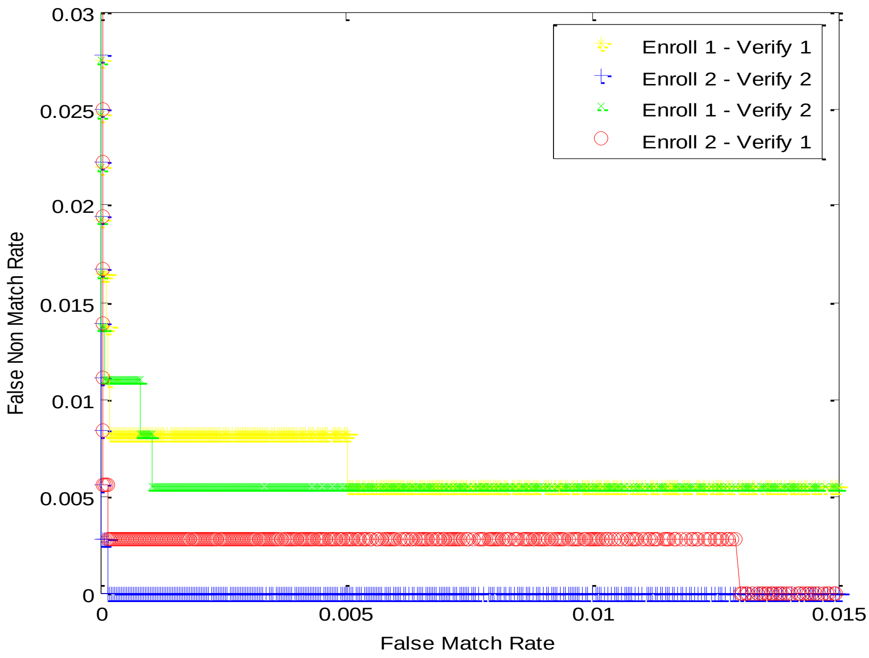
For each of the possible combinations of enrollment product and verification product, there were several Detection Error Trade-off Curves that could be computed. There were four natural subsets of the data:

1. Enroll during Visit 1, Verify during Visit 1
2. Enroll during Visit 2, Verify during Visit 2
3. Enroll during Visit 1, Verify during Visit 2
4. Enroll during Visit 2, Verify during Visit 1

Although the final performance interoperability results are based on data sets 3 and 4 only, since these are the datasets that have a time difference between enroll and verify, it is worthwhile to compare the results from these data sets to the results from data sets 1 and 2. This allows an examination of whether or not the time difference of a mean of 16 days between enroll and verify for data sets 3 and 4 really did result in a decrease in performance relative to the time difference of under an hour between enroll and verify for data sets 1 and 2.

The following figure shows a DET curve for enrollment on Product I and verification on Product I using the metric of two finger transactional FNMR on the y-axis and two finger transactional FMR on the x-axis.

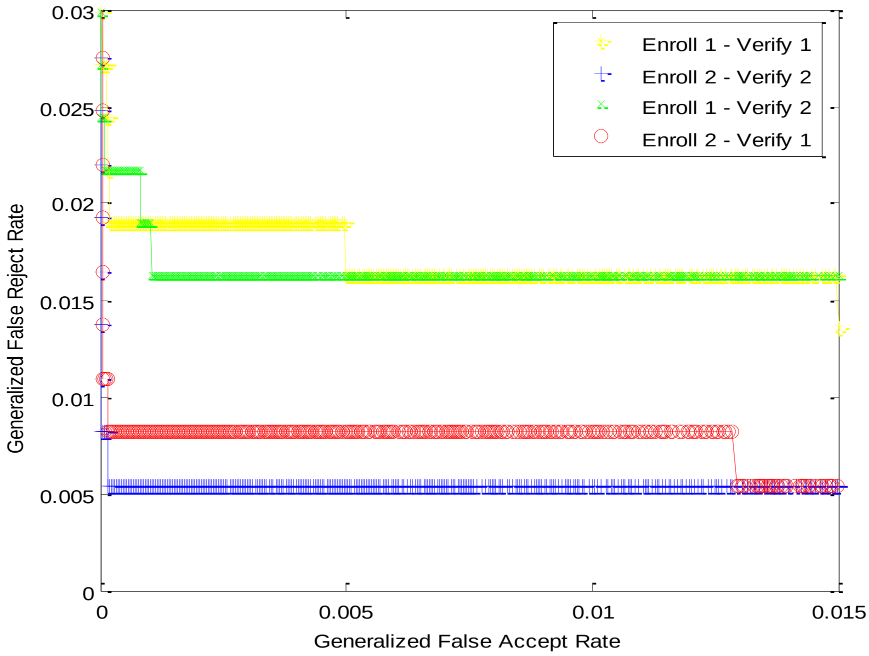
Figure 6. Enroll I Verify I – FNMR Versus FMR



The FNMR values are quantized because of the limited number of genuine transactions (approximately 368 complete two finger transactions in most cases) in each of these four data sub-sets for a specific enroll product – verify product combination. This is true for every entry in the performance interoperability matrices, each of which was computed from approximately 736 two finger genuine transactions from data sets 3 and 4. The FMR values are not significantly quantized, as the exhaustive offline testing supported a very large number of imposter transactions (typically around 130,000 per enroll sensor – verify sensor combination). Subject to the limits of the quantization, it appears that there is no significant difference between the first two data sets and the second two data sets, suggesting that other factors were more important in this experiment than the time between the enrollment and verification.

In order to properly consider the effects of FTA and FTE, a second set of DET curves were generated using the metrics of two finger transactional GFRR on the y-axis versus two finger transactional GFAR on the x-axis.

Figure 7. Enroll I Verify I – GFRR Versus GFAR



As expected, the GFRR curves on this figure have higher error rates than the FNMR curves on the previous figure. There also appears to be a potentially significant difference between the curves that used enrollment on Visit 1 and the curves that used enrollment on Visit 2. It may be that the increased familiarity of users with the fingerprint devices on their second visit (after providing a minimum of 162 fingerprint presentations during the first visit) and the improved support received from the testing staff (due to their increased experience) during the second visit, may have resulted in better enrollments during that visit. There is still, however, no significant difference attributable to the time lag between enrollment and verification. The values from this figure are more representative of real world performance than those in the previous figure and thus the final performance interoperability matrix uses the mean of the GFRR from the Enroll 2 – Verify 1 curve and the Enroll 1 – Verify 2 curve, measured at GFAR = 0.01.

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

4.3 Performance Interoperability Matrices

The performance interoperability matrices show the GFRR at a 1% GFAR for all possible combinations of enroll 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) must 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. 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 enroll product, verify product combination. Therefore the operating thresholds are not constant for each column.

The first important interoperability matrix is therefore the one that shows the results only for the three previously qualified products. In ISBIT-3, these are denoted as A0, B and C. Note that the GFRR results are shown as percentages and although two decimal places are shown, only one decimal place is significant.

Table 3. Original Three Interoperable Products, GFRR at GFAR = 1%

	A0	B	C	Mean	
A0	3.00	2.06	2.06	2.37	Enroll Product
B	5.23	0.95	1.89	2.69	
C	2.87	0.41	0.27	1.18	
Mean	3.70	1.14	1.41	2.08	
	Verify Product				

In ISBIT-2 these three products obtained a mean FRR of 0.6%, whereas in this case they obtained a mean GFRR of 2.1%. Admittedly, there was one failure to enroll in the ISBIT-1 data set that was simply ignored, but this is not enough to justify the difference. It is more likely due to a combination of environmental effects (the colder Ottawa winter drying out the skin of test subjects in comparison to the tropical waters being sailed by the Crystal Harmony in ISBIT-1 which tended to result in relatively moist skin) and the natural variation in tests caused by different test crew, different test staff, etc. Product A0 was particularly affected by this difference between the tests and its performance was substantially decreased from ISBIT-2. The vendor of product A0 had known since the later stages of ISBIT-1 that there was a bug in the quality analysis module they had shipped to ILO in their ISBIT-1 software. This meant that many images were rejected, particularly if the finger was dry and had areas with poor ridge definition. Since the test protocol did not allow for changes to acquisition algorithms during ISBIT-1 and since ISBIT-2 was an offline test that did not acquire new images, they had no chance to change this. After being added to the approved products list, they had decided to accept the error being part of the product. For ISBIT-3, however, they agreed with ILO on a compromise. They would submit a new version of their software with the image acquisition error fixed. Their image acquisition software followed the recommendation outlined in the ISBIT API Specification document (See Annex A) and returned the best image it could get, even if it determined the quality was too low, so there was no need to acquire different images with the two different products. The revised product could simply be used in the offline testing with the images obtained using the original software. This meant that there was little extra effort to test the revised version of the product, and the vendor agreed that if it proved to be superior, they would ensure that no versions of the original software were deployed in SID systems, and the original product could be removed from the approved products list.

The new version of product A0 was designated as product A1. All of the online feedback used product A0, but the offline testing was performed with both products A0 and A1. The correction of the error in the quality analysis module meant that product A1 had far fewer failure to enroll and failure to acquire errors, resulting in the table shown below.

Table 4. Interoperable Products With Bug Fixed, GFRR at GFAR = 1%

Enroll Product \ Verify Product	A1	B	C	Mean
A1	1.49	1.77	1.09	1.45
B	2.45	0.95	1.89	1.76
C	0.69	0.41	0.27	0.46
Mean	1.54	1.04	1.08	1.22

Since every entry in the table featuring product A1 was an improvement over those featuring A0, and since the mean GFRR was reduced to 1.2%, it was clear that the error in the original product did have a significant effect in the environmental conditions of this test, just as the vendor had predicted. Therefore the recommendations of this report include replacing product A0 with product A1 in the list of interoperable products. From this point on, the list of approved products is considered to be A1, B and C and all other products are required to be interoperable with them.

Having accepted A1 as the replacement for A0 in the existing list of approved products, the next important result is the performance interoperability matrix for all of the tested products. A0 is ignored since it is just the previous version of A1 with a known bug. This leaves nine products to be included in the following table.

Table 5. All Products Interoperability Matrix, GFRR at GFAR = 1%

Enroll Product \ Verify Product	A1	B	C	D	E	F	G	H	I	Mean
A1	1.49	1.77	1.09	0.82	0.54	0.81	0.95	0.82	1.08	1.04
B	2.45	0.95	1.89	0.82	0.68	2.03	0.95	2.03	2.30	1.57
C	0.69	0.41	0.27	0.27	0.27	1.24	0.27	0.27	0.27	0.44
D	0.82	0.68	0.95	0.54	0.54	1.49	0.54	0.54	0.68	0.76
E	0.68	0.82	0.81	0.54	0.14	1.49	0.82	0.82	1.22	0.82
F	1.09	1.37	0.82	0.54	0.54	0.82	0.82	0.54	0.68	0.80
G	1.36	1.50	1.35	0.68	0.55	1.90	0.54	0.82	0.95	1.07
H	0.68	0.27	0.54	0.27	0.27	1.23	0.41	0.27	0.68	0.51
I	1.23	2.47	1.09	0.82	0.82	1.78	1.09	0.82	1.23	1.26
Mean	1.16	1.14	0.98	0.59	0.48	1.42	0.71	0.77	1.01	0.92

In the table above, the native performance of each product (that is the performance when the same product is used for enrollment and verification) is highlighted in gray and lies on the diagonal of the matrix. Native performance varies from 0.14% to 1.49%. The interoperable performance makes up the off-diagonal elements of this performance interoperability matrix and varies from 0.27% to 2.47%. The mean of the entire matrix is 0.92%. This means that all of the products can interoperate at the required ILO performance threshold of GFRR < 1% at GFAR = 1% and the addition of the new products into the performance interoperability matrix substantially improves performance over using the three previously approved products, which had a mean GFRR of 1.22%, even after the bug in product A0 was fixed. Even the maximum value of GFRR in the performance interoperability matrix of 2.47% is not in any significant way worse than the maximum value recorded among the three approved products of 2.45%. This leads to the conclusion that all of the six new tested products should be added to the approved products list.

Further analysis of this table leads to the conclusion that some products (such as C and H) are better at generating interoperable biometric information records (BIRs) and some products (such as D and E) are better at matching against BIRs 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 some interesting insights.

Table 6. All Products Interoperability Matrix, FNMR at FMR = 1%

Enroll Product \ Verify Product	A1	B	C	D	E	F	G	H	I	Mean
A1	1.23	1.50	0.27	0.54	0.27	0.27	0.68	0.55	0.54	0.65
B	2.45	0.95	1.22	0.82	0.68	0.54	0.95	2.03	2.04	1.30
C	0.42	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
D	0.82	0.68	0.27	0.54	0.54	0.00	0.54	0.54	0.41	0.48
E	0.68	0.82	0.14	0.54	0.14	0.00	0.82	0.82	0.95	0.54
F	0.82	1.10	0.00	0.27	0.27	0.00	0.55	0.27	0.27	0.39
G	1.36	1.50	0.68	0.68	0.55	0.41	0.54	0.82	0.68	0.80
H	0.41	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.41	0.11
I	0.42	1.66	0.00	0.00	0.00	0.00	0.28	0.00	0.41	0.31
Mean	0.95	0.93	0.29	0.38	0.27	0.14	0.50	0.56	0.64	0.52

As expected, the overall performance is improved, with the mean of the entire performance interoperability matrix dropping to 0.52%. 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 product F. In the previous table, products D and E were best at interoperable matching, as determined by the mean of their respective columns in the interoperability matrix. Now they are joined by products C and F, with F having the best performance of any product at interoperable matching. In the previous table, however, when the effects of FTE and FTA are included, product F had the worst performance of all matchers. It is possible that Product F simply had a very strict quality analysis algorithm and rejected a lot of reasonable images, but it is also possible that Product F simply had an excellent algorithm paired with a mismatched sensor. This shows the importance of measuring sensor effects. A product can have excellent algorithms, but coupling it with a specific sensor can move it from best to worst performance in a given test.

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 draft standard ISO 19795-4 Biometric Performance Testing and Reporting Part 4 – Performance and Interoperability Testing of Interchange Formats.

There are approximately 736 genuine transactions representing approximately 4416 single finger matches used to compute the GFRR value in each position in the performance interoperability matrix. The average of these values is 0.92%. Being conservative, a one sided 95%, confidence interval based on the assumption of 736 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:

$$Z_{95\%,\ 1\ \text{sided}} = 1.645\ (p * (1-p) / 736)^{0.5} = 0.006 = 0.6\%$$

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 81 entries in the performance interoperability matrix, we can run the same calculation for the mean, using 81 * 736 = 59616 independent measurements and using the exact value of p = 0.0092. This gives a 95%, one sided confidence interval of 0.06%, suggesting that there is only a 5% chance that the mean GFRR would be larger than 0.98%. 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, etc. do not substantially affect the result.

Another option for exploring the confidence in the data is to break it back into the two separate data sets representing enrollment during the first visit and verification during the second visit (E1V2) versus enrollment during the second visit and verification during the first visit (E2V1). Although the natural variability in GFRR at 1% GFAR in each of these data sets will be larger than for the combined data set (due to only having half of the independent 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 7. All Products Interoperability Matrix for E1V2 Only, GFRR at GFAR = 1%

Enroll Product \ Verify Product	A1	B	C	D	E	F	G	H	I	Mean
A1	1.36	2.17	0.82	0.54	0.54	0.54	0.82	0.54	0.82	0.91
B	1.90	1.36	2.17	1.09	0.82	2.16	1.09	2.14	1.90	1.62
C	0.82	0.82	0.54	0.54	0.54	1.36	0.54	0.54	0.54	0.69
D	0.82	0.54	0.81	0.54	0.54	1.36	0.54	0.54	0.54	0.69
E	0.54	0.54	0.54	0.54	0.00	1.36	0.54	0.54	1.36	0.66
F	1.09	1.09	0.54	0.54	0.54	1.09	0.54	0.54	0.81	0.75
G	1.36	1.09	0.81	0.82	0.27	1.89	0.54	0.82	1.08	0.96
H	0.82	0.54	0.54	0.54	0.54	1.36	0.82	0.54	0.82	0.72
I	1.09	2.45	1.09	1.09	1.09	1.90	1.36	1.09	1.63	1.42
Mean	1.09	1.18	0.87	0.69	0.54	1.45	0.75	0.81	1.06	0.94

Table 8. All Products Interoperability Matrix for E2V1 Only, GFRR at GFAR = 1%

Enroll Product \ VerifyProduct	A1	B	C	D	E	F	G	H	I	Mean
A1	1.63	1.37	1.36	1.09	0.55	1.08	1.09	1.09	1.35	1.18
B	2.99	0.55	1.62	0.54	0.55	1.90	0.82	1.91	2.71	1.51
C	0.56	0.00	0.00	0.00	0.00	1.13	0.00	0.00	0.00	0.19
D	0.82	0.82	1.08	0.54	0.55	1.63	0.54	0.55	0.81	0.82
E	0.82	1.09	1.08	0.54	0.27	1.63	1.09	1.09	1.08	0.97
F	1.09	1.65	1.09	0.54	0.54	0.54	1.09	0.54	0.54	0.85
G	1.36	1.91	1.89	0.54	0.82	1.90	0.54	0.82	0.81	1.18
H	0.55	0.00	0.55	0.00	0.00	1.09	0.00	0.00	0.55	0.30
I	1.37	2.49	1.10	0.54	0.54	1.65	0.82	0.54	0.82	1.10
Mean	1.24	1.10	1.09	0.48	0.42	1.40	0.67	0.73	0.97	0.90

Although there are individual elements in the two interoperability matrices that are different, the means of the two matrices are 0.94% and 0.90% respectively. It is impossible to determine if this difference is because of a systematic change between E1V2 and E2V1, such as the increased familiarity of the test subjects during the second visit leading to better enrollments, or whether it is simply part of the normal variation between two independent sets of measurements. In either case, the number is well within the 95% confidence interval predicted using the normal approximation and the z-test. By looking at the differences in individual elements of the interoperability matrices, we can also evaluate whether a 95% confidence interval of 0.6% for each individual entry was reasonable.

Table 9. Interoperability Matrix Difference, E1V2 – E2V1

Enroll Product \ Verify Product	A1	B	C	D	E	F	G	H	I	Mean
A1	-0.27	0.81	-0.54	-0.54	0.00	-0.54	-0.27	-0.55	-0.54	-0.27
B	-1.09	0.81	0.54	0.54	0.27	0.27	0.27	0.23	-0.81	0.11
C	0.25	0.82	0.54	0.54	0.54	0.23	0.54	0.54	0.54	0.51
D	0.00	-0.28	-0.27	0.00	0.00	-0.27	0.00	0.00	-0.27	-0.12
E	-0.27	-0.55	-0.54	0.00	-0.27	-0.27	-0.54	-0.55	0.27	-0.30
F	-0.01	-0.56	-0.55	0.00	0.00	0.54	-0.55	0.00	0.27	-0.09
G	0.00	-0.83	-1.08	0.27	-0.55	0.00	0.00	0.00	0.27	-0.21
H	0.27	0.54	0.00	0.54	0.54	0.27	0.82	0.54	0.27	0.42
I	-0.29	-0.04	-0.01	0.54	0.54	0.25	0.54	0.54	0.81	0.32
Mean	-0.16	0.08	-0.21	0.21	0.12	0.05	0.09	0.08	0.09	0.04

The Table above shows the differences between the performance interoperability matrix computed using E1V2 and that computed using E2V1. If E2V1 was used as the baseline, and we assumed a 95%, 1 sided confidence interval of 0.006, as computed above for the complete test, we would expect 4 elements of this matrix to exceed 0.6%. In fact, only 3 elements exceed this threshold. Similarly, if E1V2 was used as the baseline, we would expect 4 elements of the matrix to be lower than –0.6%. Once again, only 3 elements exceed this threshold. Since the uncertainty for an experiment based on only half the samples is actually larger than the uncertainty for the complete experiment by a factor of $2^{0.5}$, this suggests that our earlier estimate of 0.6% as the 95% upper uncertainty bound on each element in the matrix may in fact be quite conservative.

Of course, this is only a very simplistic analysis, but it does give some degree of confidence that the individual elements in the interoperability matrix are reasonably reliable. The mean is even more reliable, ensuring that this group of nine interoperable products is likely to meet the ILO performance interoperability target of a mean GFRR of 1% if tested again, and hopefully when deployed in real world environments.

5 Interoperable Product Combinations

ISBIT-3 was designed to determine whether or not there were any additional products, beyond the three on the published ILO approved products list, which could conform to the conformance, performance and interoperability requirements of the ILO. Since all products evaluated were conformant and since the mean of the performance interoperability matrix containing the three previously approved products (one of which was revised to fix a bug) and the six new products was 0.92%, it was clear that all nine 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 nine 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 nine products at various sizes of sub-group.

Table 10. 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**
3	A1,B,C	1.22%	2.45%
4	A1,B,C,E	0.93%	2.45%
5	A1,B,C,E,H	0.84%	2.45%
6	A1,B,C,D,E,H	0.77%	2.45%
7	A1,B,C,D,E,G,H	0.79%	2.45%
8	A1,B,C,D,E,G,H,I	0.86%	2.47%

For further comparison purposes, the table below provides the same analysis if there was no requirement to include the three previously approved products. It shows that in many cases, the new products actually perform better. This is not surprising, given that the approved products were submitted for testing in ISBIT-1, over 18 months prior to the submission deadline for ISBIT-3. This serves as a reminder that fingerprint technology is still advancing and vendors may wish to periodically submit their most up to date products for possible inclusion in the approved products list.

Table 11. 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**
3	C,E,H	0.41%	0.82%
4	C,D,E,H	0.46%	0.95%
5	C,D,E,G,H	0.53%	1.35%
6	C,D,E,G,H,I	0.64%	1.35%
7	A1,C,D,E,G,H,I	0.72%	1.49%
8	A1,C,D,E,F,G,H,I	0.80%	1.90%

6 Conclusions and Recommendations

Based on the results of the ISBIT-3 test, the following conclusions can be made:

1. Product A0 should be withdrawn from the approved products list and replaced by product A1.

2. All six of the new products (D, E, F, G, H and I) meet the ILO requirements for conformance, performance and interoperability and should be added to the approved products list.

3. Sensor effects are very important and estimation of real world performance really requires that products be tested as paired sensors and algorithms. Algorithm only performance can be very different from total product performance.

4. The overall performance of fingerprint technology using standardized data records to achieve interoperability has improved significantly since ISBIT-1 in 2004. The ILO approved products list and other similar lists of interoperable products should be periodically refreshed as fingerprint technology improves.

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 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: Performance and Interoperability Testing of Data Interchange Formats

8 Acknowledgments

The authors of this report would like to gratefully acknowledge the editorial assistance of Christopher Craig, Cynthia Musselman, and James Wilson in preparing this report. Particular appreciation is also due to the experts group from ISO/IEC JTC 1 SC 37 for providing external review and feedback on this report.

Glossary

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

administrator
person performing the testing or enrollment

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

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

Biometric Interchange Record (BIR)
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)

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

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

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

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

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 enroll rate (FTE)
proportion of the population for whom the system fails to complete the enrollment process

NOTE The observed failure to enroll rate is measured on test crew enrollments. The predicted/expected failure to enroll 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 enrollment 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 enrollment 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 BIR) on the part of a user simulated during offline testing resulting in a verification decision

NOTE If a BIR 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 enrollment and matching separately from data capture

NOTE 1 Collecting a database of samples for offline enrollment 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 enrollment 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 enrollment and for calculation of matching scores at later attempts. The images/signals collected can be used immediately either for an online enrollment, verification, or identification attempt, or may be stored and used later for offline enrollment, 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 enrollment 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 enrollment 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 enrollment or verification

NOTE There are two types of transactions: enrollment sequence, resulting in an enrollment or a failure to enroll; 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: ISBIT–3 API Specification

This is the document that was provided to the vendors prior to the test so that they could provide products that could be easily integrated into the test harness.

API Specification

ILO SID Biometric Interoperability Test

November 2005

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 BIRs 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 BIRs 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 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.

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 enrollment 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 enrollment 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 enrollment 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 6.

`purpose` (input): A value of zero (0) will indicate a capture for the purpose of enrollment, 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 enroll both primary and secondary fingers as an SID-0002 BIR 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 BIR 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 BIR shall be 'unenrolled'. Return values -1, -2, and -3 will indicate that the enrollment 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 6.

fingerSecondary (input): A value from 1 to 10 corresponding to a valid finger position from SID-0002 or ANSI/NIST-ITL 1-2000, table 6.

`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 BIR in bytes.

`bir` (output): SID-0002 BIR 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 Enroll Primary
-2 Failed to Enroll Secondary
-3 Failed to Enroll 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 BIR 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 BIR. 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 BIR in bytes.

`bir` (input): SID-0002 BIR 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 BIR. A value of zero indicates that the intermediate template should be matched with the primary template of the SID-0002 BIR.

`score` (output): A similarity score resulting from the comparison of the intermediate template with the primary or secondary template of the SID-0002 BIR. 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 BIR
-3	Failed to Process Intermediate Template and BIR

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 – 3 Methodology Description for Vendors

Methodology

ILO SID Biometric Interoperability Test

November 2005

Bion BIOMETRICS

B.1 Terms and Definitions

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

administrator
person performing the testing or Enrollment

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

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

Biometric Identification Record (BIR)
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)

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

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

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

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

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 enroll rate (FTE)
proportion of the population for whom the system fails to complete the enrollment process

NOTE The observed failure to enroll rate is measured on test crew enrollments. The predicted/expected failure to enroll 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 enrollment 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 enrollment 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 BIR) on the part of a user simulated during offline testing resulting in a verification decision

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

observer
test staff member recording test data or monitoring the crew

offline testing
execution of enrollment and matching separately from data capture

NOTE 1 Collecting a database of samples for offline enrollment 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 enrollment 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 enrollment and for calculation of matching scores at later attempts. The images/signals collected can be used immediately either for an online enrollment, verification, or identification attempt, or may be stored and used later for offline enrollment, 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 enrollment 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 enrollment 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 enrollment or verification

NOTE There are two types of transactions: enrollment sequence, resulting in an enrollment or a failure to enroll; 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 19795-2 Biometric Performance Testing and Reporting – Part 4: Performance and Interoperability Testing of Data Interchange Formats

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

Only two of the seven products, A and F, achieved the ILO targets for both native and interoperable performance, and so it became apparent that interoperability using the standard might not be as simple as had been anticipated. A follow-on study was commissioned to investigate what had caused the problems in interoperability. 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. These products are listed in a document provided to the ILO Governing Body that can be found, at the time of writing, at http://www.ilo.org/public/english/standards/relm/gb/docs/gb292/pdf/gb-16-2-ad.pdf.

The ILO plans to conduct 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.

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: enrollment and verification.

During enrollment, a test subject may attempt to enroll 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 may attempt to match their primary or secondary finger with a BIR 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 testing lab will be tested for conformance to ensure that they can produce and read fingerprint templates in the form of the Biometric Interchange Records (BIRs) defined in Annex B of SID-0002. If they are conformant, then they will be integrated into unified test control 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 150 people will enroll on each system and attempt to verify multiple times on each system against BIRs 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 enroll on one system and verify on another to be explored for both genuine and impostor distributions. The original set of three products will then have their average interoperability performance score normalized to that of the previous test and any other products that can be added to the subset without raising that average interoperability error rate will then be considered as meeting the relevant requirements of the Seafarers' Identity Documents Convention (No. 185).

Performing third-party, independent testing of biometric products from several vendors for both enrollment 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.

To this end, the ILO SID Biometric Interoperability Test is partitioned into two phases:

Phase 1 – Integration and Conformance: Performed with a very limited test crew to determine if the biometric product was conformant to the required BIR format and to perform initial interoperability tests. Only those products that satisfactorily complete the first phase of the test are included in the second phase of the test.

Phase 2 – Performance and Interoperability: Conducted with a test crew comprised of approximately 150 test subjects. Each test subject will make two visits to the test lab for biometric enrollment, verification, and related data collection. Online and offline matching will help to determine if the BIRs generated by one biometric product can be used to verify test subjects using the same and other vendors' biometric products. This test also determines if each biometric product can verify test subjects using BIRs generated by other vendors' biometric products.

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 three members: an Experimenter, Administrator, and Observer. 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 Observer reads prepared scenario scripts to each test subject explaining the purpose and conduct of the test, and guides each test subject through the enrollment and verification visits. The Administrator ensures that the test system functions properly, enters test subjects' information, records test details such as finger used, enrollment results, verification results, and generates performance reports. The Observer and Administrator verbally corroborate results to mitigate any potential data collection errors.

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 enrollment 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 will be 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.

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.

The demographics (nationality, job type, age, etc) of the test crew will be recorded and as individuals sign up, running totals will be maintained. Reasonable efforts will be made to attain a similar demographic distribution as that of the test crew of the first ILO Biometric Test Campaign onboard the Crystal Harmony and this may require that some potential test subjects will eventually be turned down because different demographics are needed. This will, of course, be balanced with the need to acquire a sufficient volunteer test crew in a timely fashion.

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 enroll 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.

For each test subject, the administrator will enter the following data into the test control software's database: test subject ID, birth year, nationality group, job group, and gender of the test subject, current temperature, and humidity at the time of each visit, and any additional comments that may be useful.

Scenario scripts will be read to each test subject at the beginning of both enrollment and verification sessions. The observer 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 and/or observer will also, whenever a test subject has problems using a biometric product, provide some 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

Great lengths will be taken to protect the personal information of the test subjects. The database will never contain names, addresses, exact birth dates or other personally identifying details. Instead, these will be maintained in a paper record that will be kept in a secure area. The paper record will also contain each person's test subject ID, and it is this number that will be the link to the data in the database. This will allow the paper record to be used to ensure that the same test subject is present for both visits and to find a particular test subject or group of test subjects if they need to be called back in for some additional testing. The test database, however, will have no way of linking the generic demographic data or the fingerprint data to specific individuals.

After the test is over the paper record will be moved to a secure off-site storage, where it can still be referenced if there is ever a request from an individual to find out what data on them has been retained. It is expected that the fingerprint images and templates will be retained 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 during both the first and second visit.

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, pending open solicitations or announcements made by the ILO. Such announcements explain the ILO's intention to create a list of biometric systems that successfully satisfy the objectives of the test for use by ratifying Members of ILO Convention No. 185. 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 believed would be most advantageous to them (for a seafaring population) for the purposes of the test.

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 is evaluated for stability and its effects on the stability of other biometric products in the test harness. With the cooperation of the vendor with the test lab, any issues related to integration and conformance (detailed in the following section) must be resolved before 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, Enrollment, Verification, and Nominal Interoperability.

B.8.1 Enrollment

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

- prompts for placement of all ten possible finger positions by name
- provides visual feedback of the fingerprint image presented to the sensor
- indicates a failure to acquire or failure to enroll 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 BIRs 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
- 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 BIRs
- 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 BIRs enrolled on each of the other biometric products. The product is considered to have passed a single interoperability test for a particular BIR 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 BIRs from products other than itself, and 50% of those other products have to successfully pass interoperability tests with the given product's BIR.

For example, with 7 total products, the BIR from Product 1 containing the right and left index fingers from a single tester will be used to attempt a successful verification on each of Products 2 through 7. Similarly, the tester will attempt a successful verification on Product 1 against the BIRs from each of Products 2 through 7. For Product 1 to be considered interoperable, at least 50% of six, or three of these six single interoperability tests must match when verification is being attempted using Product 1 and at least three must match when verification is being attempted on Products 2 through 7 against the

BIR from Product 1. The process would be repeated for a small group of well-habituated testers and the average number of passed tests for that group of testers 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 and non-native 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. Interoperability testing is representative of a real port-based seafarer verification scenario.

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 Enrollment

Test subjects are enrolled on each biometric product during both visits in accordance with the requirements stated in ILO SID-0002. During Enrollment, a test subject will make two enrollment attempts to enroll 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 team will try to enroll a fingerprint from 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 enroll 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 BIRs will be stored in a secure database for subsequent online and offline verifications.

B.9.2 Online Verification

After the enrollment 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 results.

The test control 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 BIR 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 enrollment BIR. If a BIR 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 calculated using the two finger (six-presentation) offline matching transactions described above. For each biometric product, the native (enroll on that product and verify on that product) performance will be evaluated to find the threshold that produces a 1% false accept rate. That threshold will then be used with the transactional verification similarity scores for all verifications using that product and the false reject rate computed for enrollment with each product and verification using the current product. This will produce one column of a two dimensional interoperability matrix where the rows represent enrollment products and the columns represent verification products. This will be repeated for every other product to build up a complete matrix, as shown below.

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

The existing interoperable products identified by the ILO will also be part of this matrix and all of the FRR values in the matrix will be divided by a scalar factor defined such that the average of the FRR values in all combinations of the existing interoperable products is 1.0%. After that, all possible subsets of the products (always including the existing interoperable products) will be considered and the largest subset that produces an interoperability matrix with an average FRR value of less than or equal to 1% will be defined as the new group of interoperable products. If more than one subset of equal size produces an interoperability matrix with average FRR less than or equal to 1%, then the subset that produces the lowest average FRR will be selected.

Annex C: Two finger transaction Based DET Curves

Figure 8. Enroll A1 Verify A1 – FNMR Versus FMR

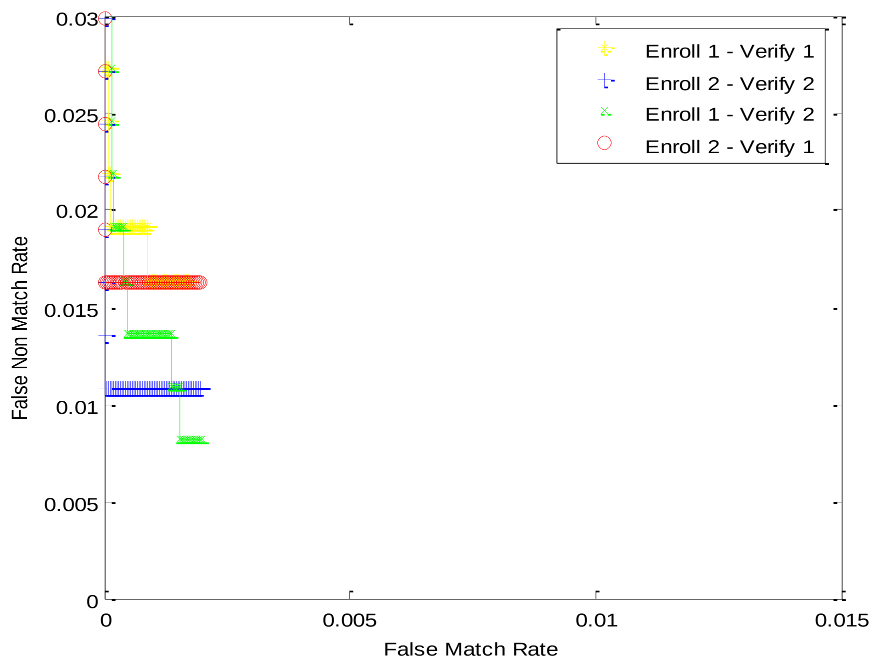


Figure 9. Enroll A1 Verify A1 – GFRR Versus GFAR

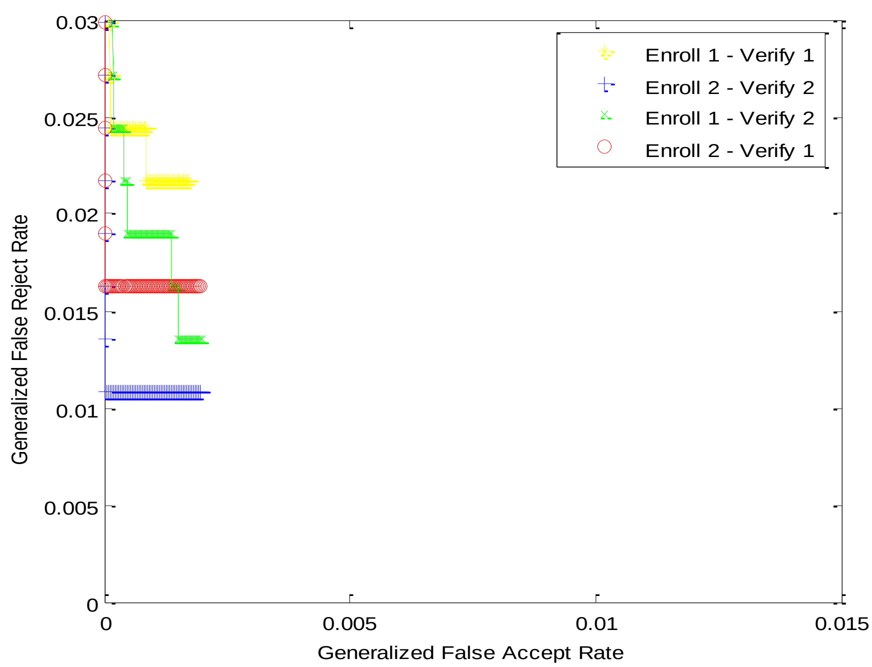


Figure 10. Enroll A1 Verify B – FNMR versus FMR

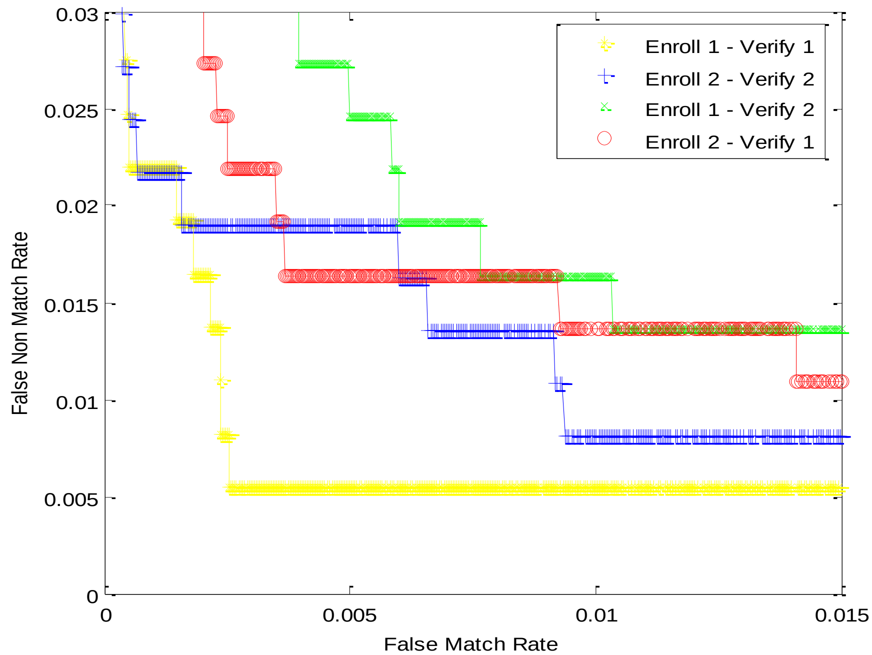


Figure 11. Enroll A1 Verify B – GFRR versus GFAR

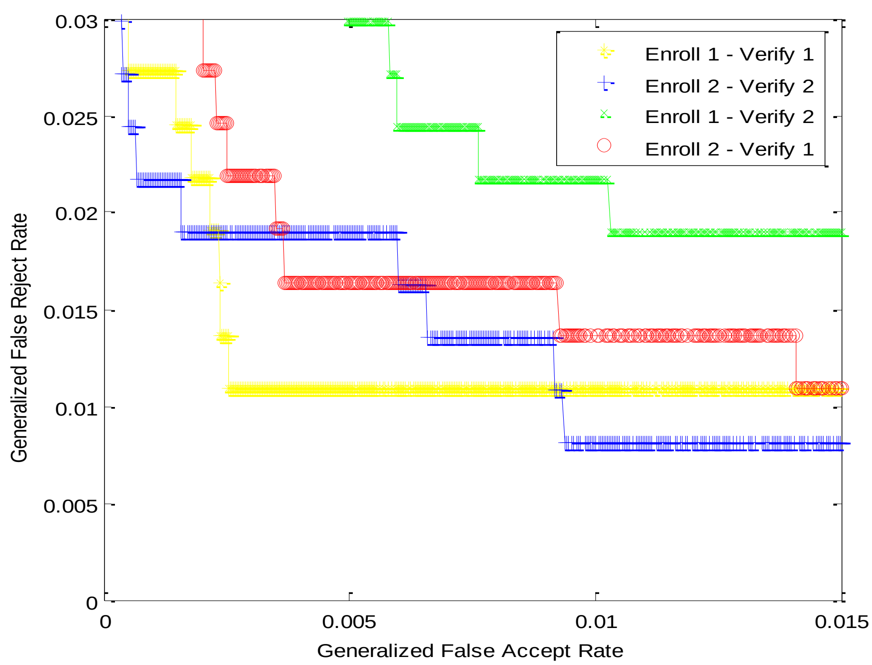


Figure 12. Enroll A1 Verify C – FNMR Versus FMR

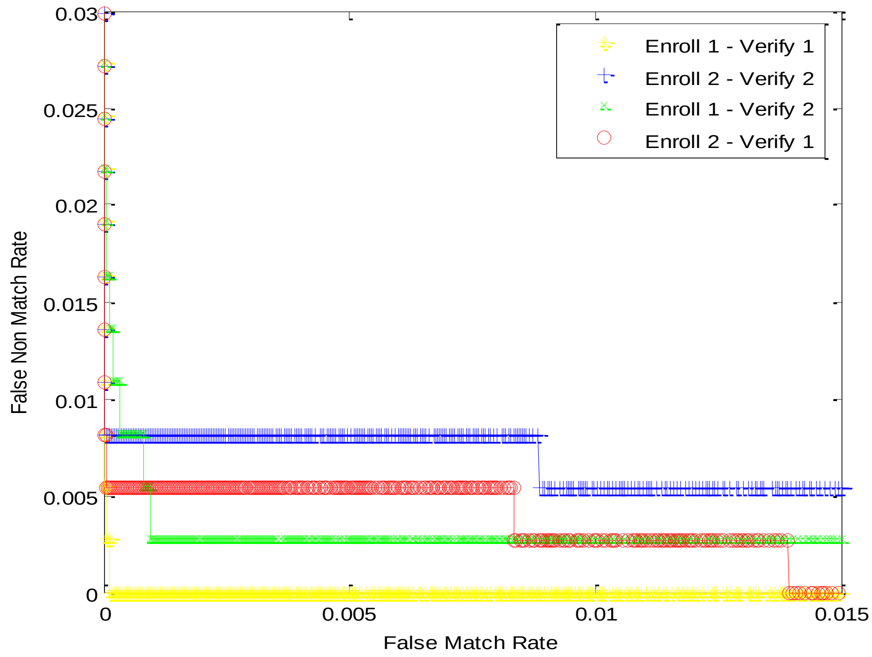


Figure 13. Enroll A1 Verify C – GFRR Versus GFAR

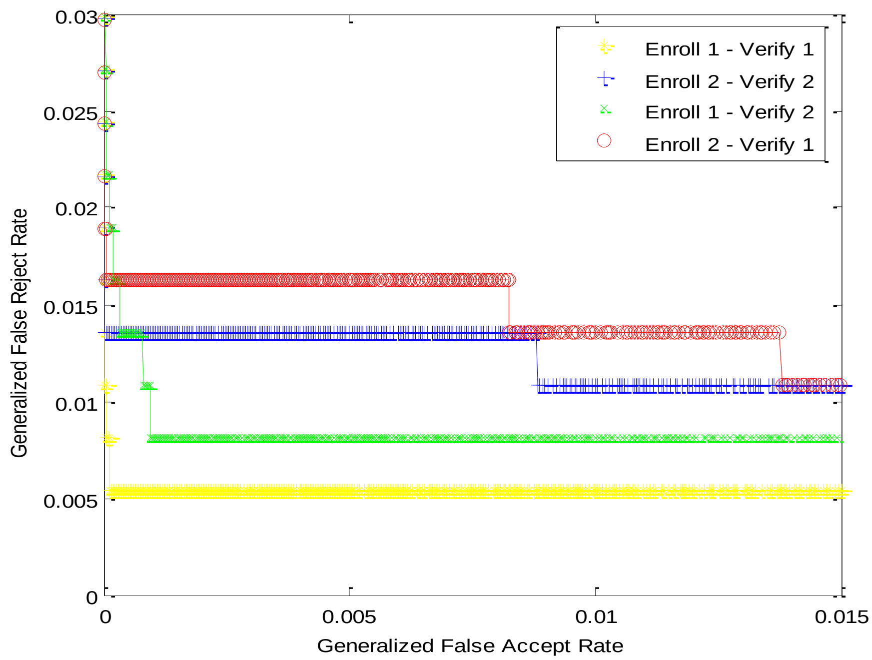


Figure 14. Enroll A1 Verify D – FNMR Versus FMR

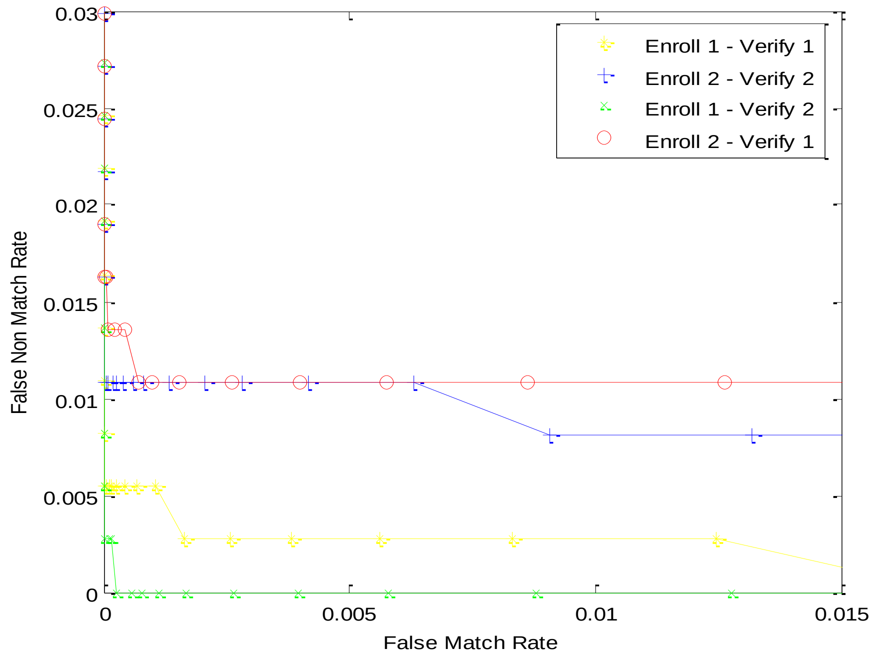


Figure 15. Enroll A1 Verify D – GFRR Versus GFAR

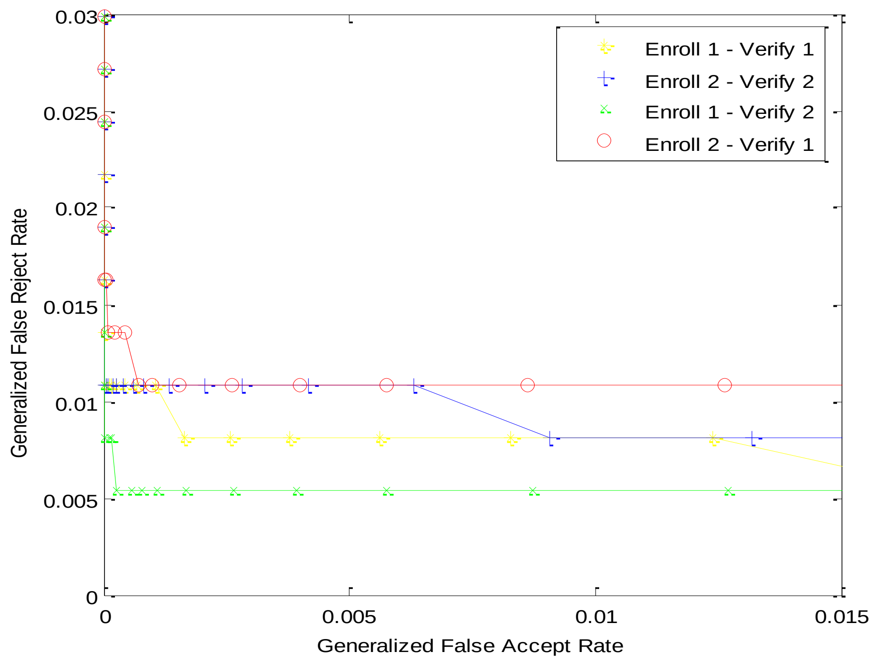


Figure 16. Enroll A1 Verify E – FNMR Versus FMR

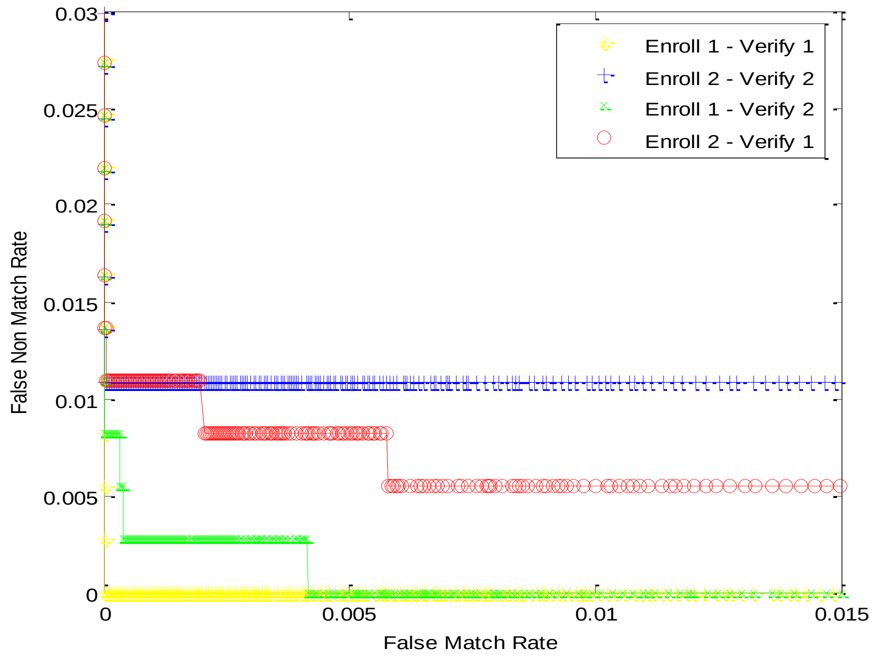


Figure 17. Enroll A1 Verify E – GFRR Versus GFAR

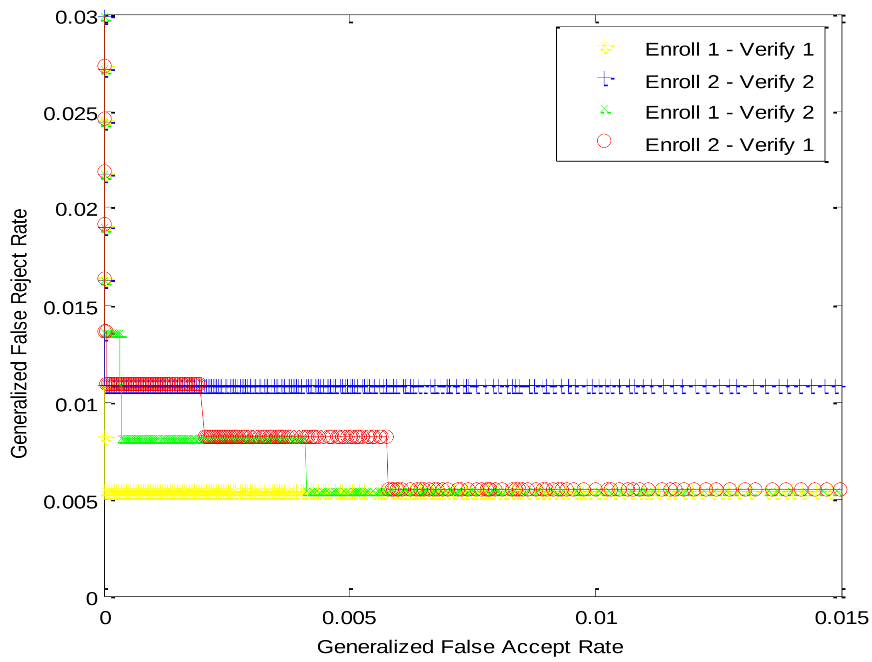


Figure 18. Enroll A1 Verify F – FNMR Versus FMR

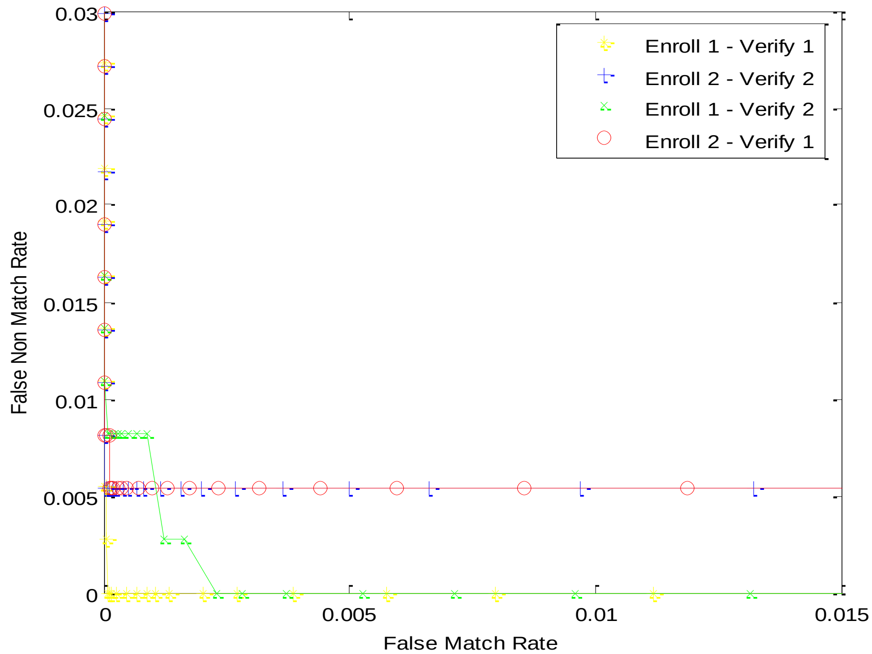


Figure 19. Enroll A1 Verify F – GFRR Versus GFAR

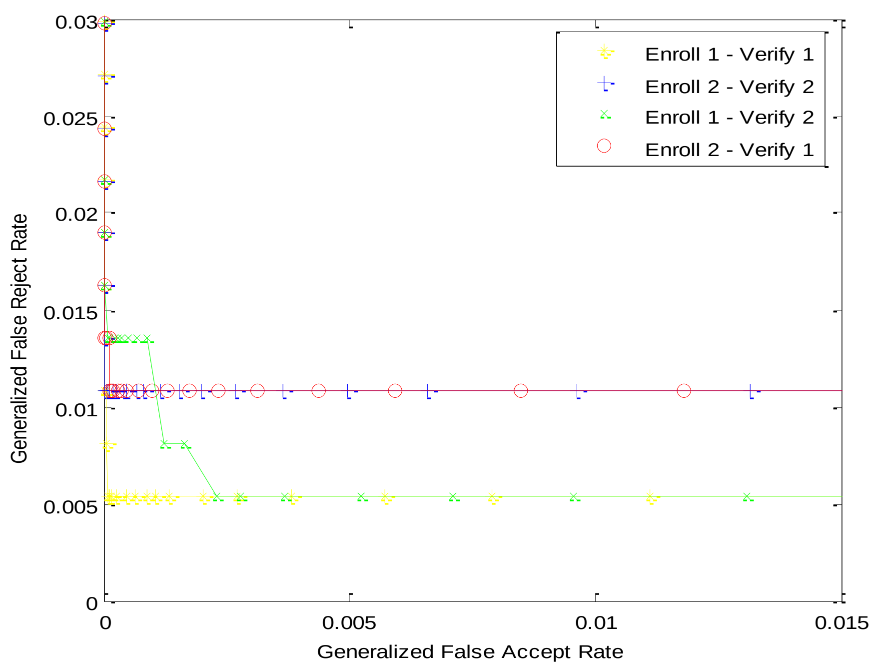


Figure 20. Enroll A1 Verify G – FNMR Versus FMR

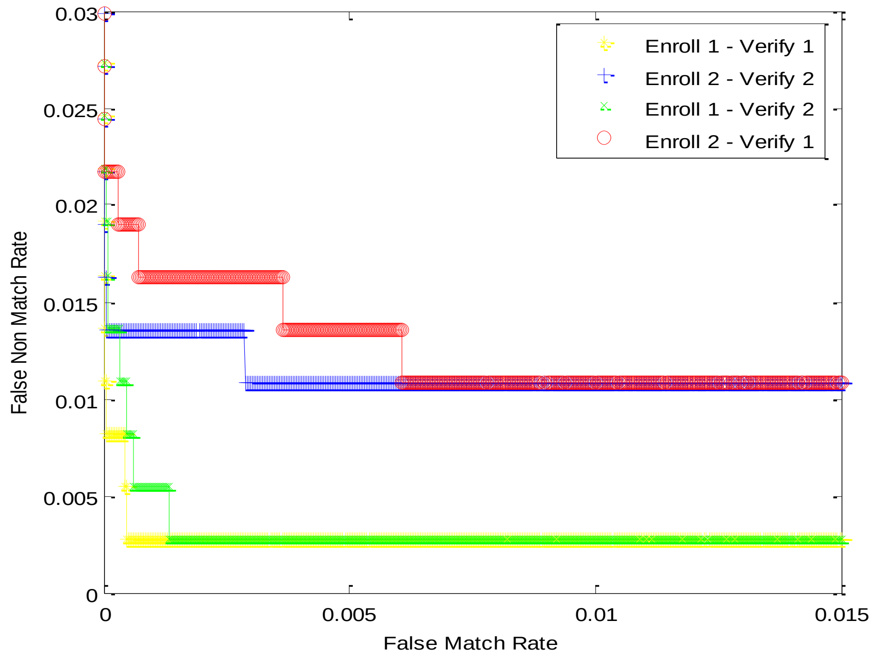


Figure 21. Enroll A1 Verify G – GFRR Versus GFAR

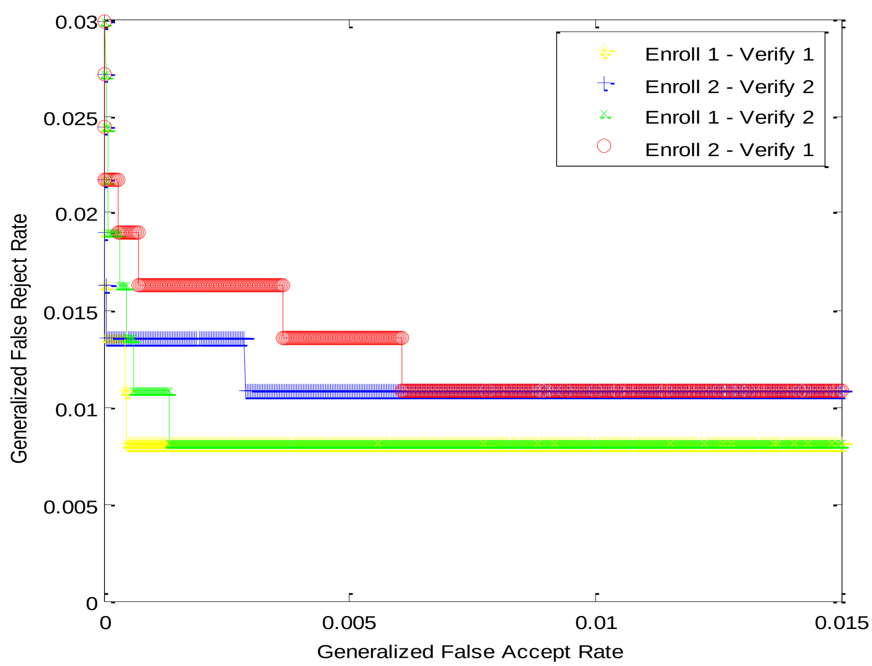


Figure 22. Enroll A1 Verify H – FNMR Versus FMR

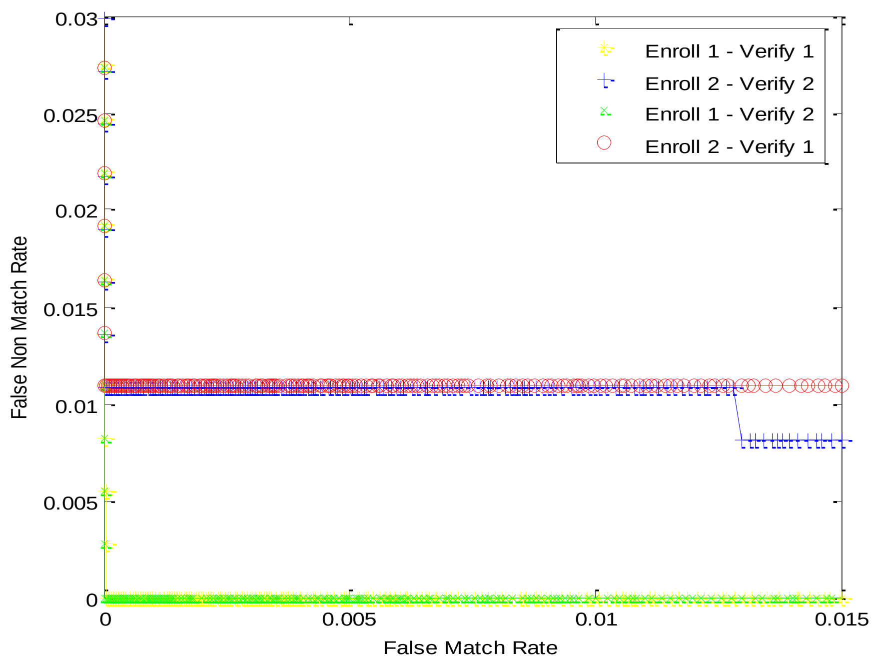


Figure 23. Enroll A1 Verify H – GFRR Versus GFAR

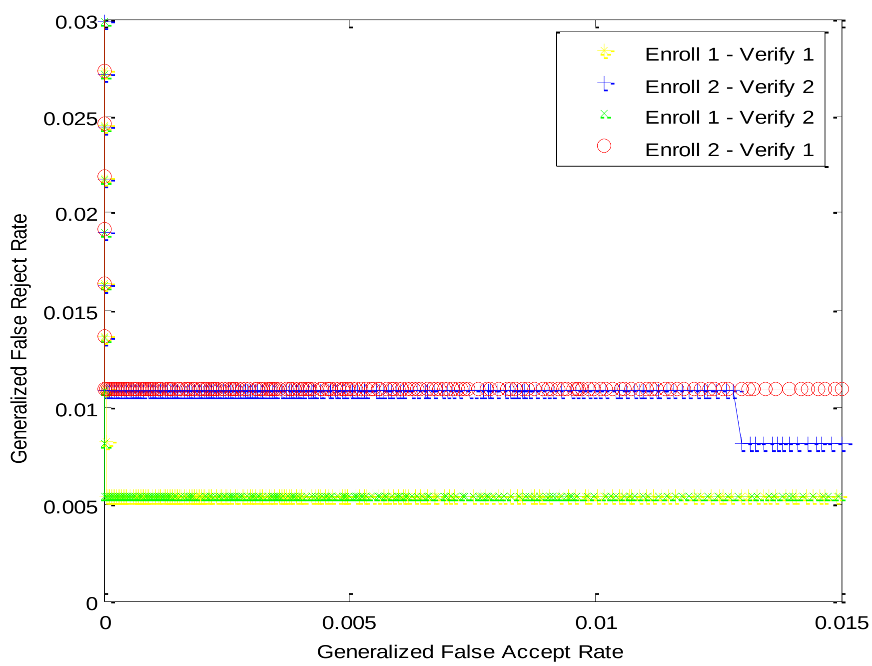


Figure 24. Enroll A1 Verify I – FNMR Versus FMR

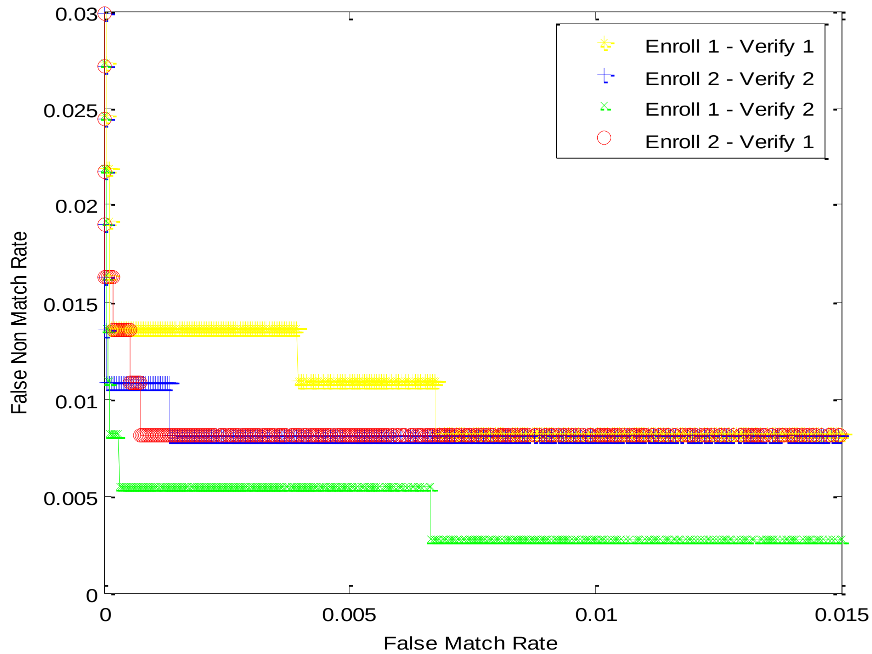


Figure 25. Enroll A1 Verify I – GFRR Versus GFR

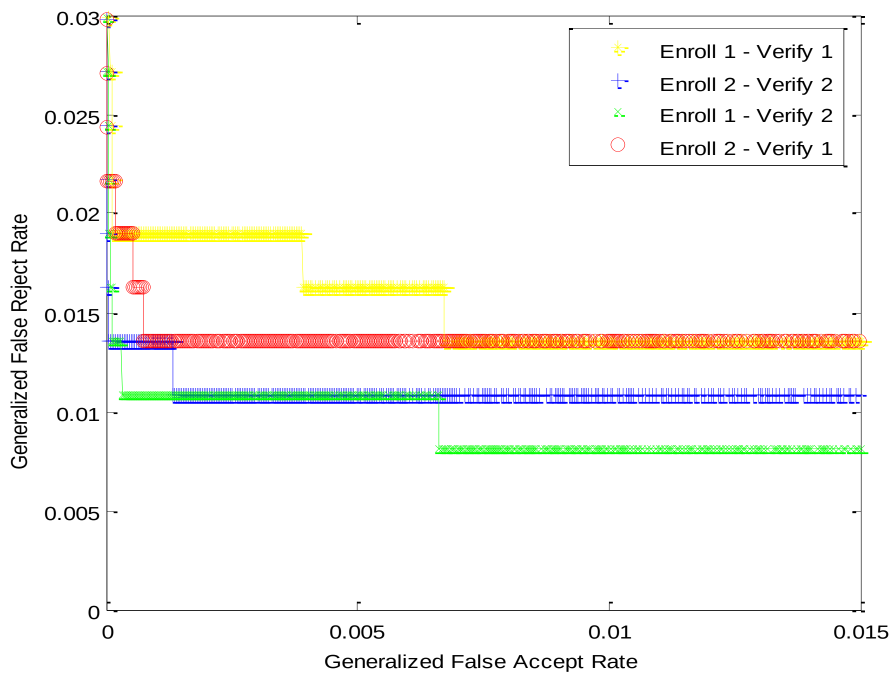


Figure 26. Enroll B Verify A1 – FNMR Versus FMR

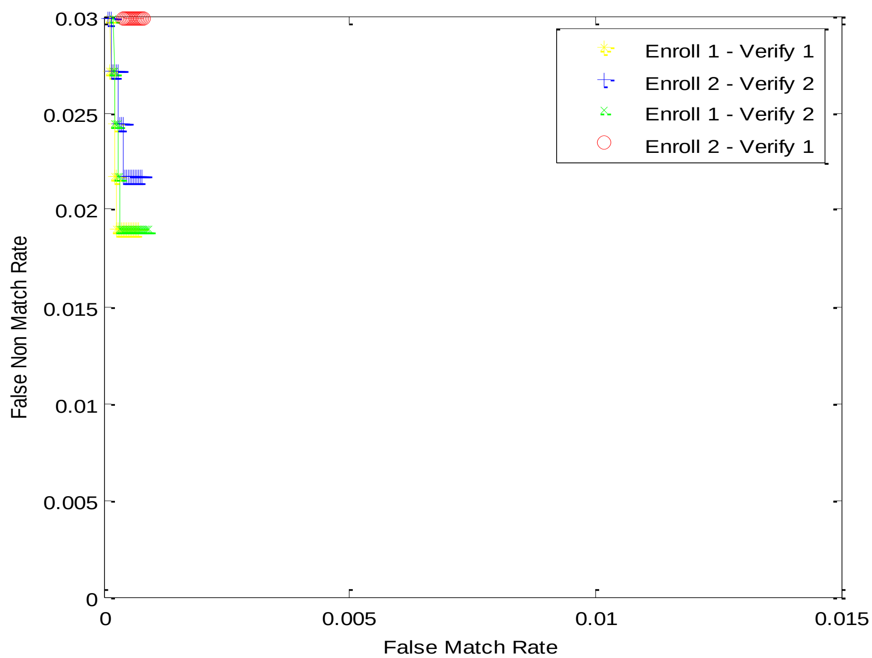


Figure 27. Enroll B Verify A1 – GFRR Versus GFAR

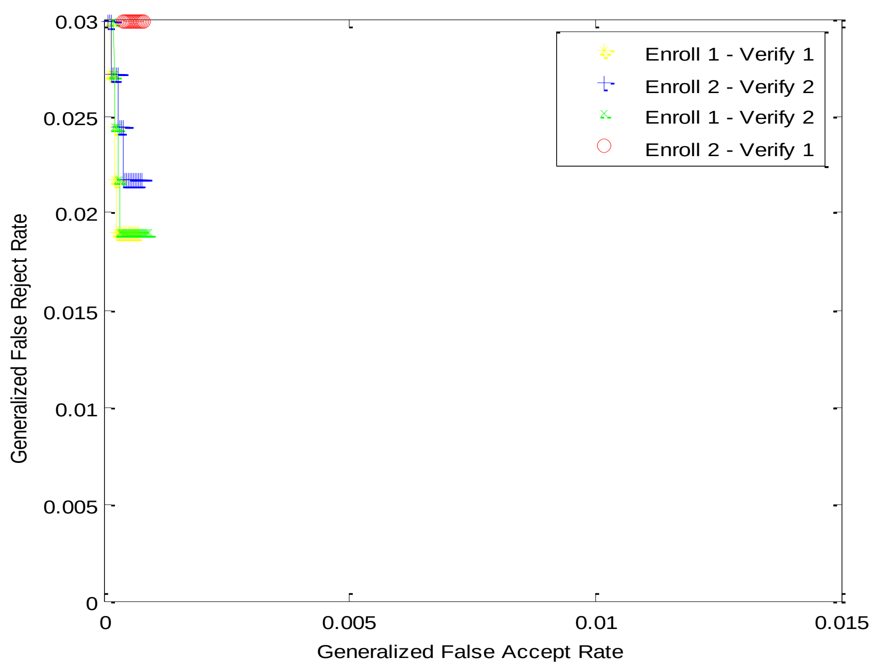


Figure 28. Enroll B Verify B – FNMR Versus FMR

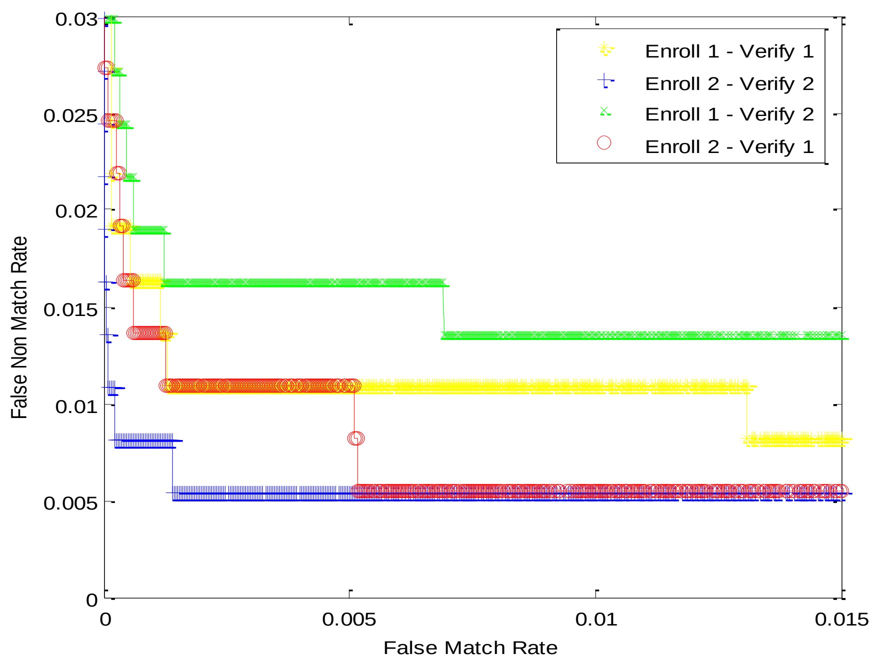


Figure 29. Enroll B Verify B – GFRR Versus GFAR

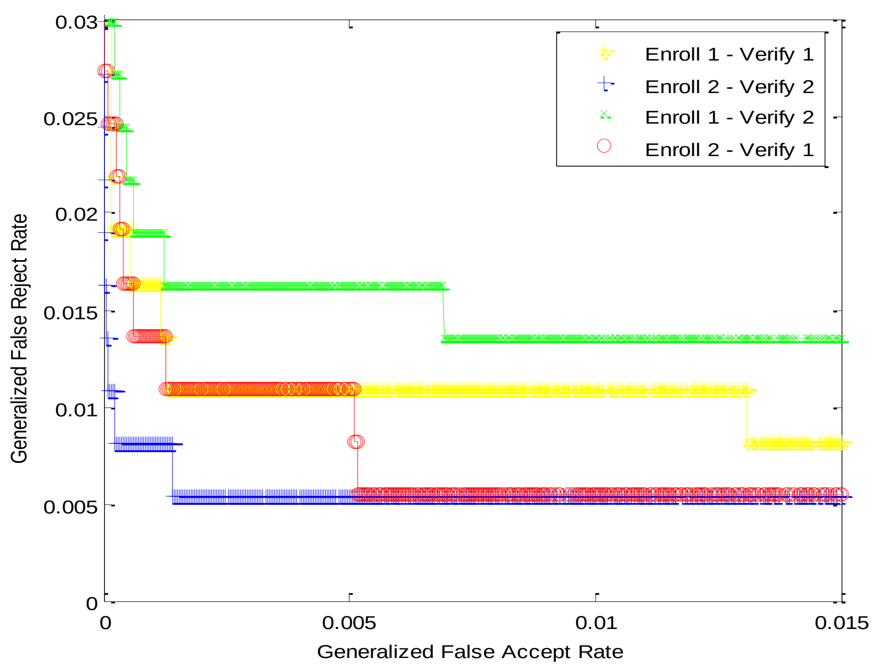


Figure 30. Enroll B Verify C – FNMR Versus FMR

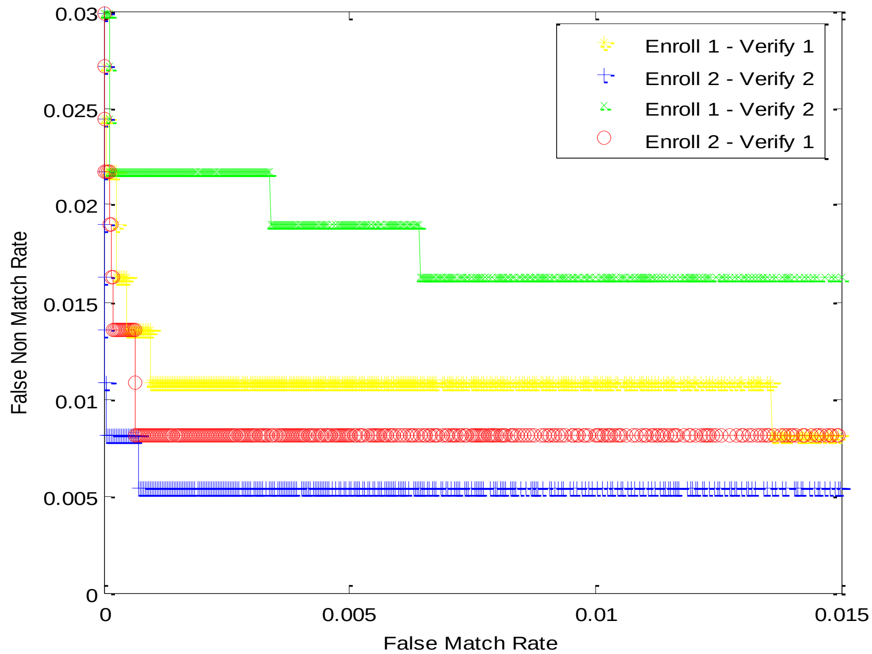


Figure 31. Enroll B Verify C – GFRR Versus GFAR

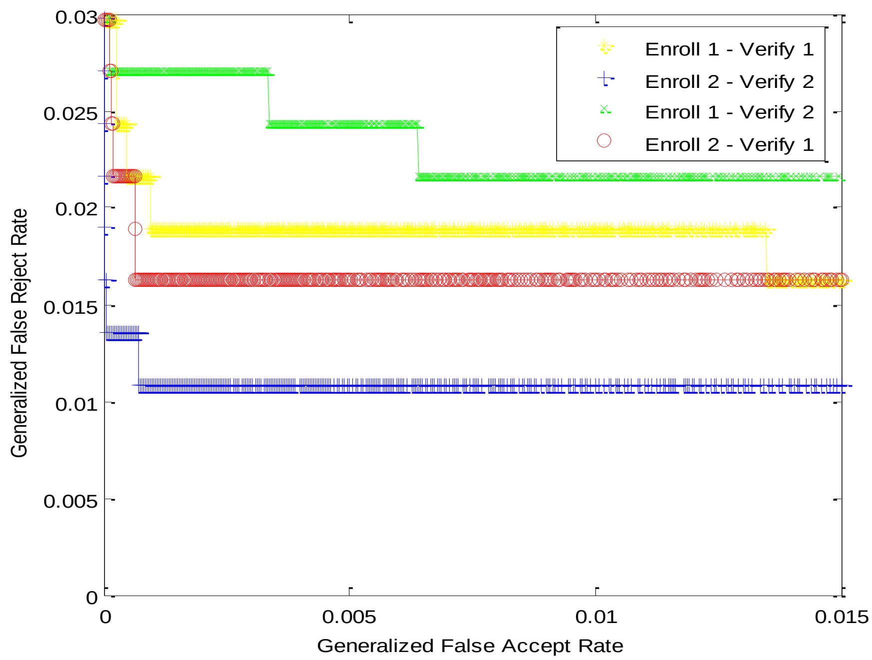


Figure 32. Enroll B Verify D – FNMR Versus FMR

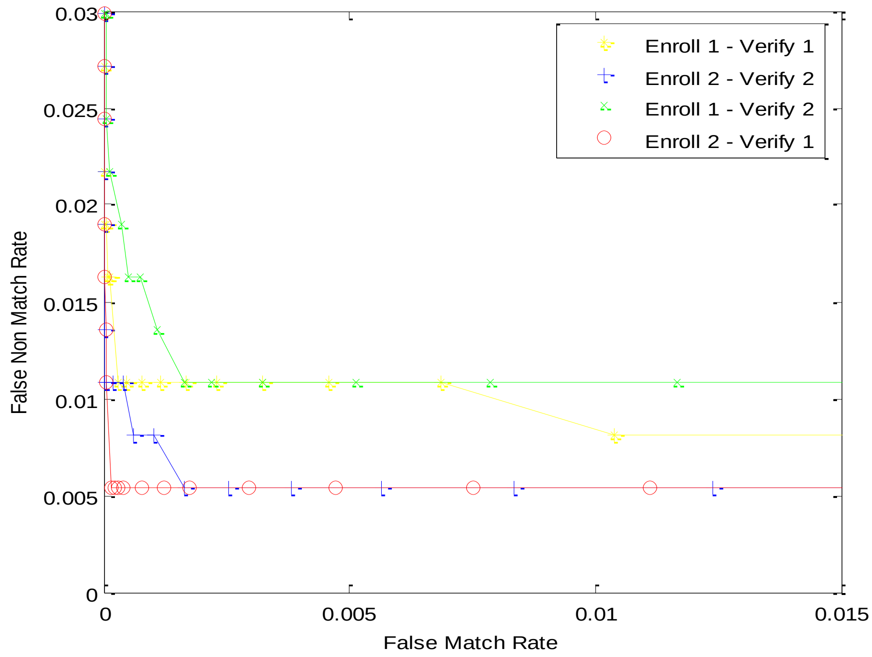


Figure 33. Enroll B Verify D – GFRR Versus GFAR

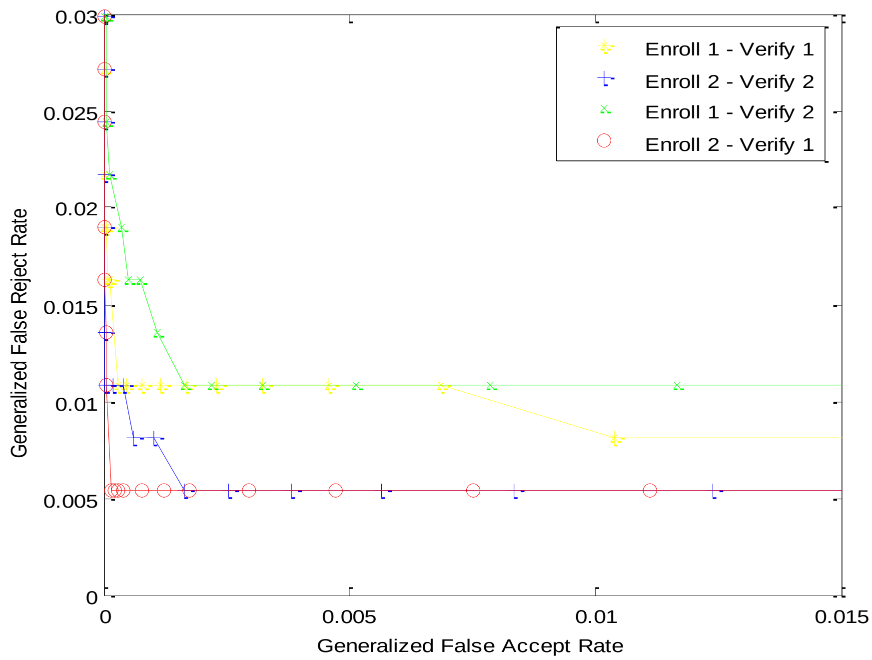


Figure 34. Enroll B Verify E – FNMR Versus FMR

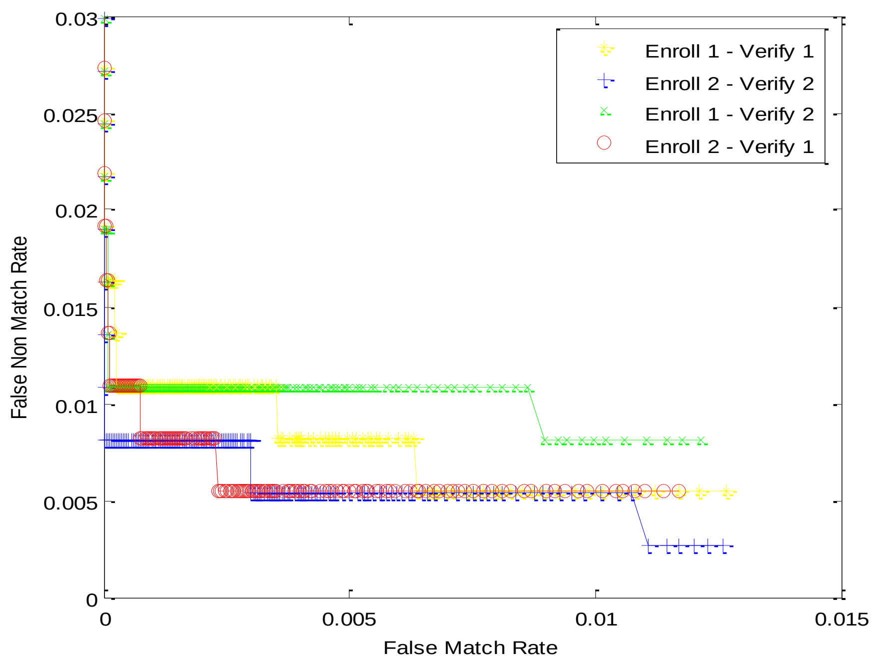


Figure 35. Enroll B Verify E – GFRR Versus GFAR

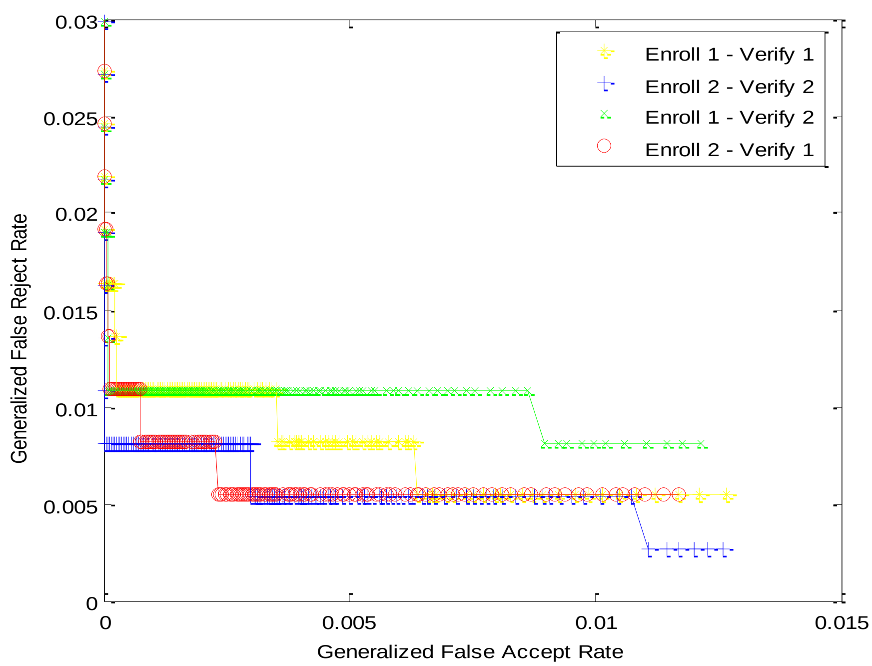


Figure 36. Enroll B Verify F – FNMR Versus FMR

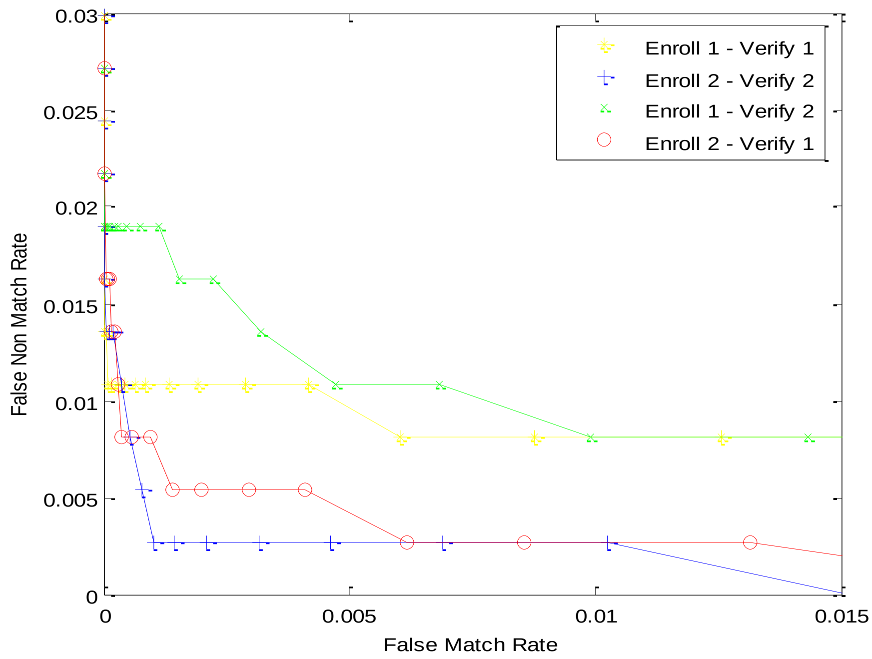


Figure 37. Enroll B Verify F – GFRR Versus GFR

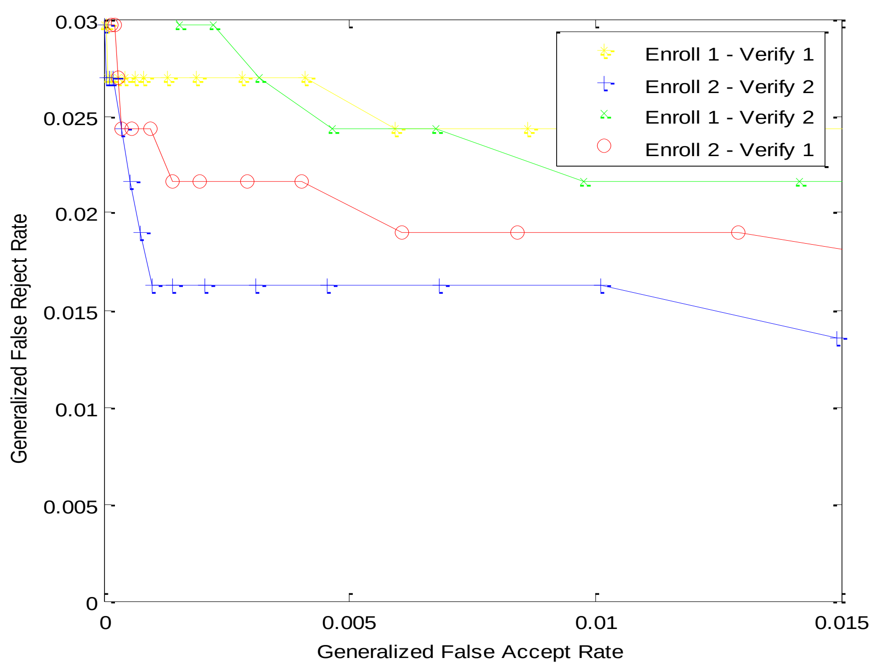


Figure 38. Enroll B Verify G – FNMR Versus FMR

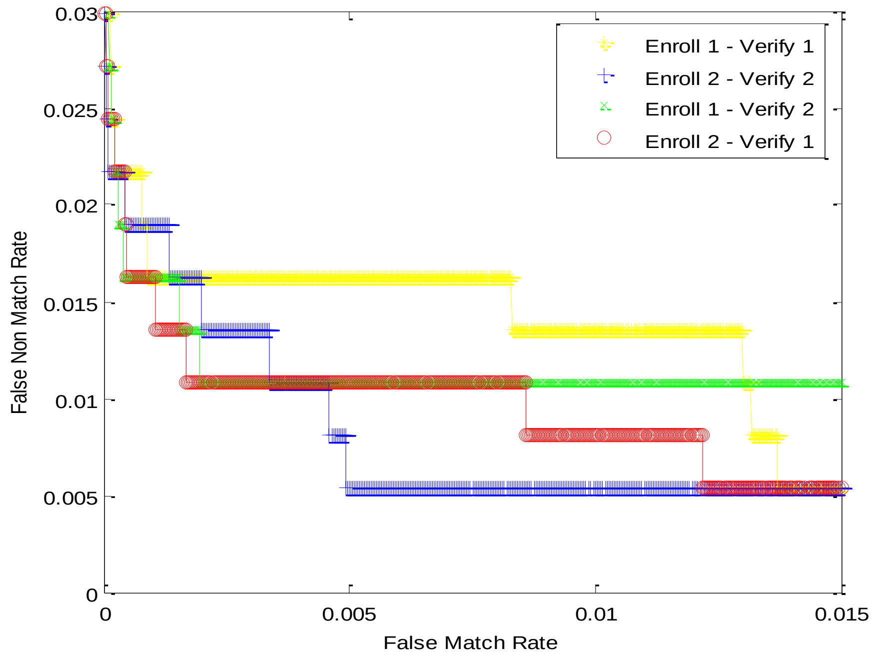


Figure 39. Enroll B Verify G – GFRR Versus GFAR

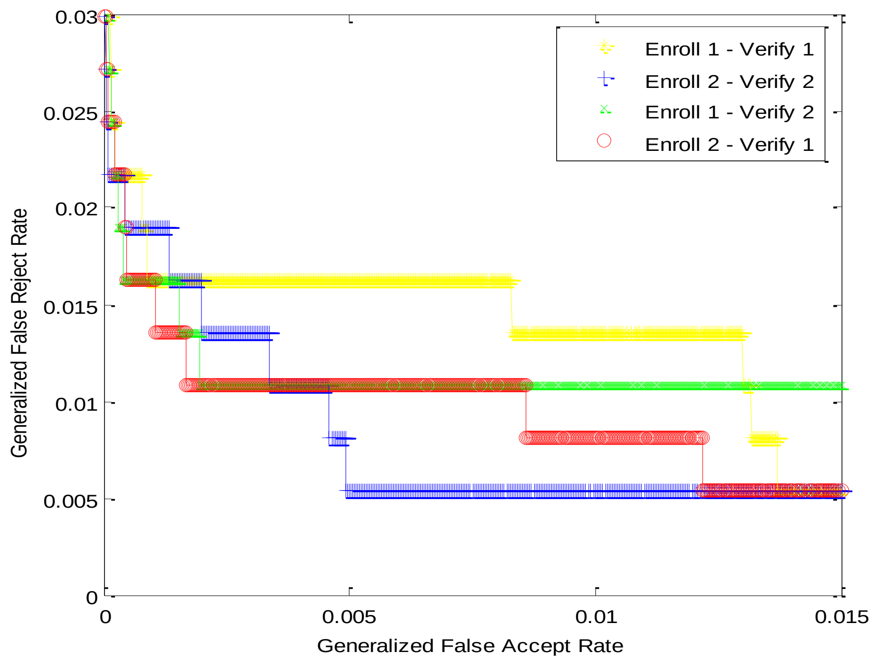


Figure 40. Enroll B Verify H – FNMR Versus FMR

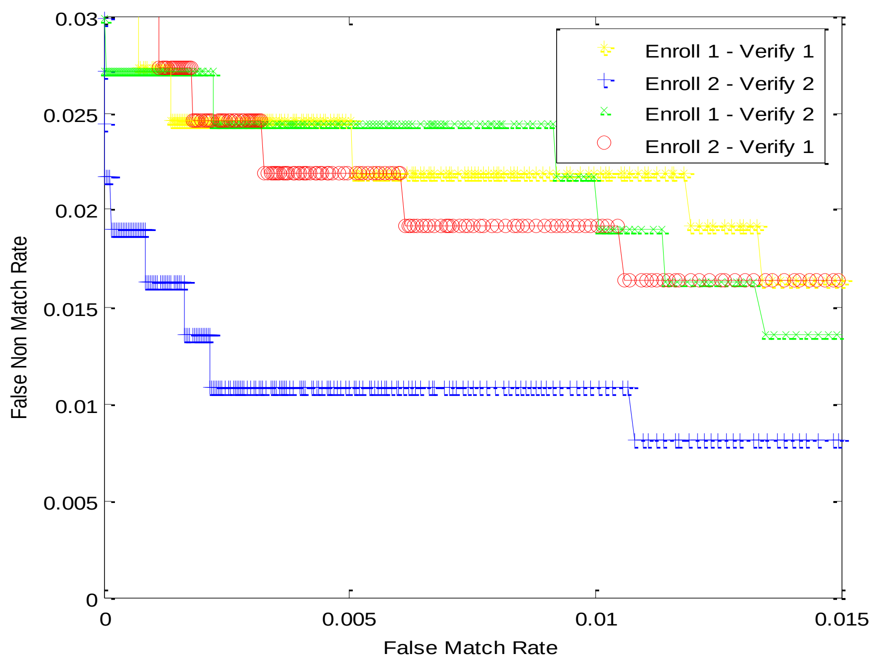


Figure 41. Enroll B Verify H – GFRR Versus GFAR

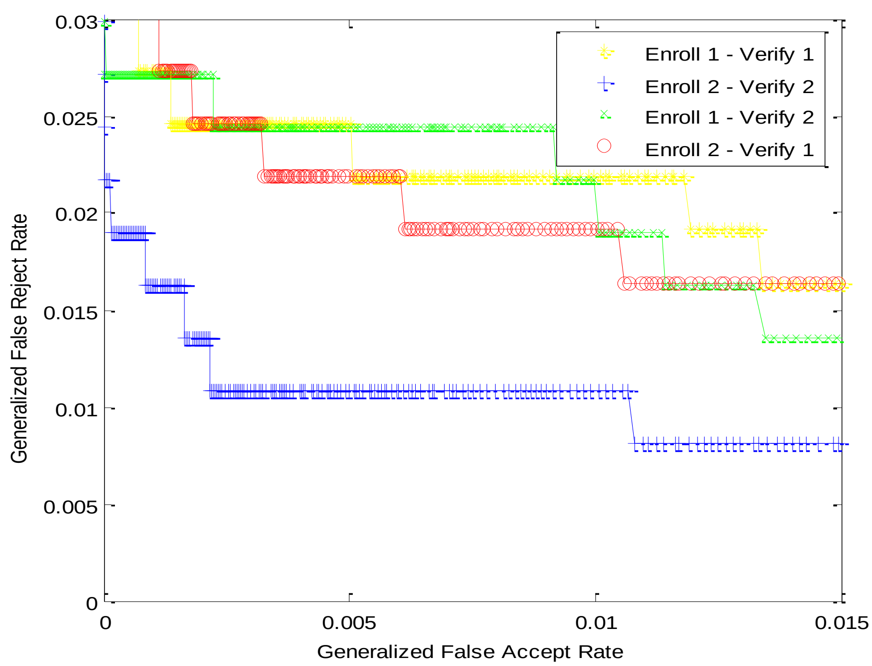


Figure 42. Enroll B Verify I – FNMR Versus FMR

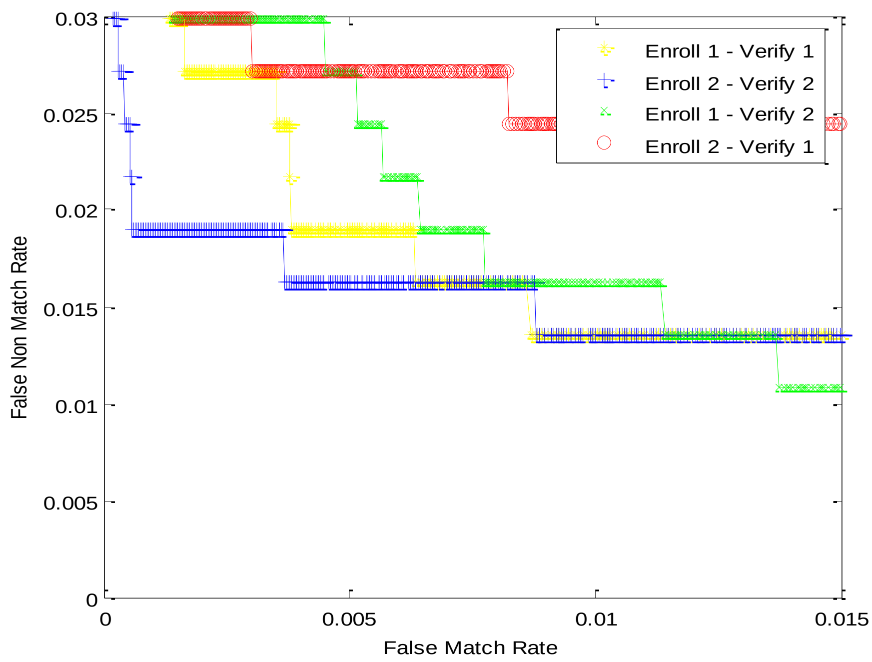


Figure 43. Enroll B Verify I – GFRR Versus GFAR

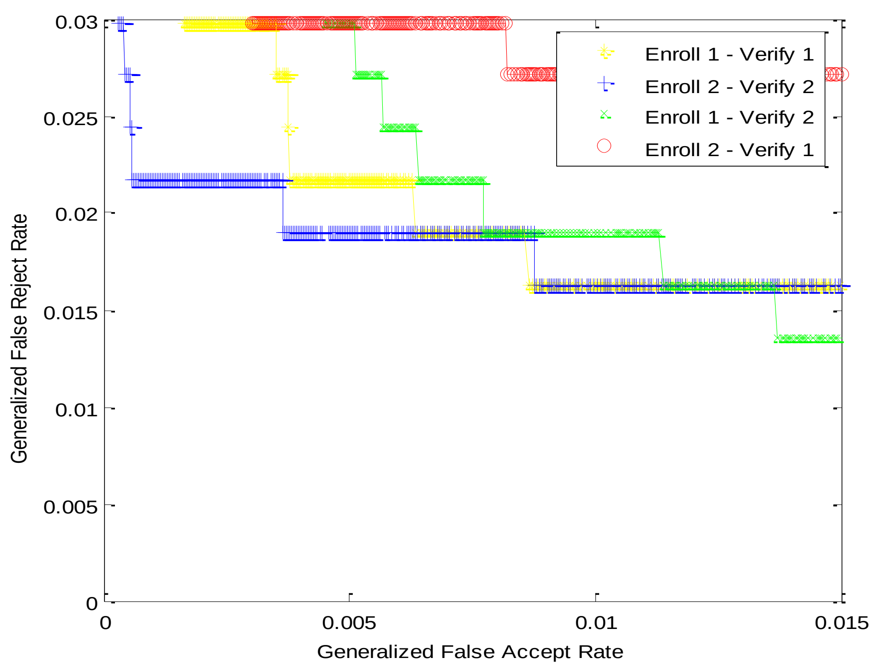


Figure 44. Enroll C Verify A1 – FNMR Versus FMR

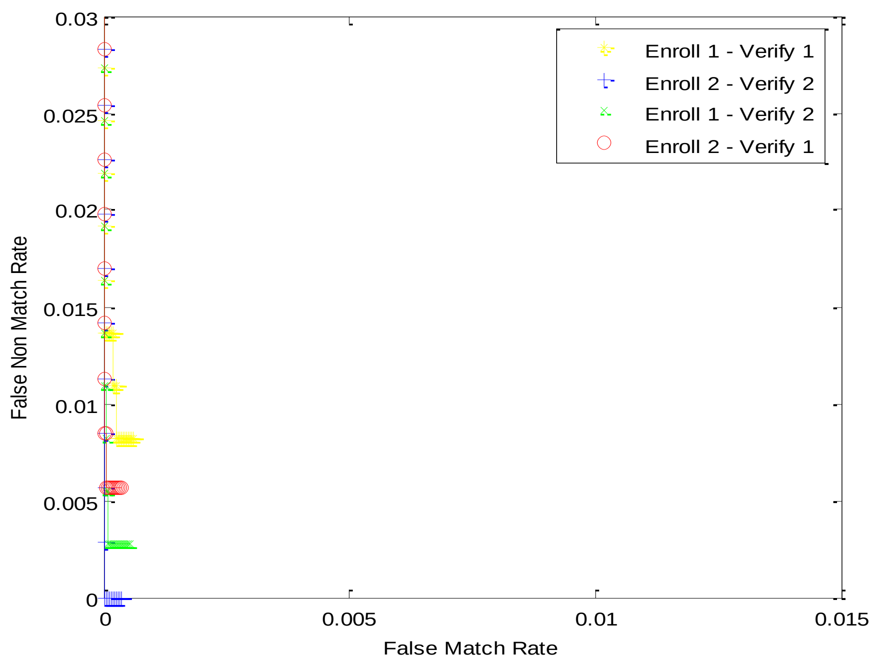


Figure 45. Enroll C Verify A1 – GFRR Versus GFAR

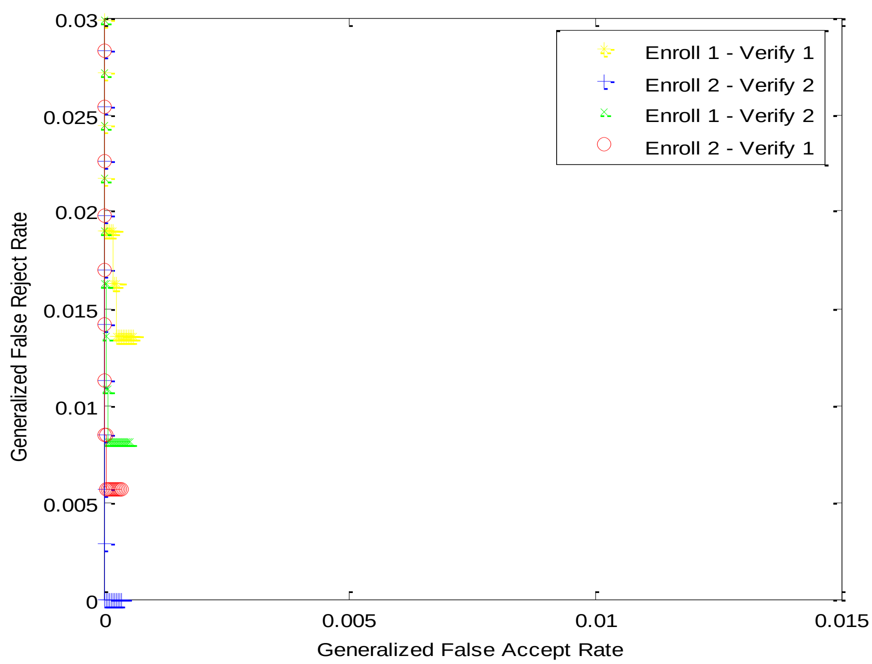


Figure 46. Enroll C Verify B – FNMR Versus FMR

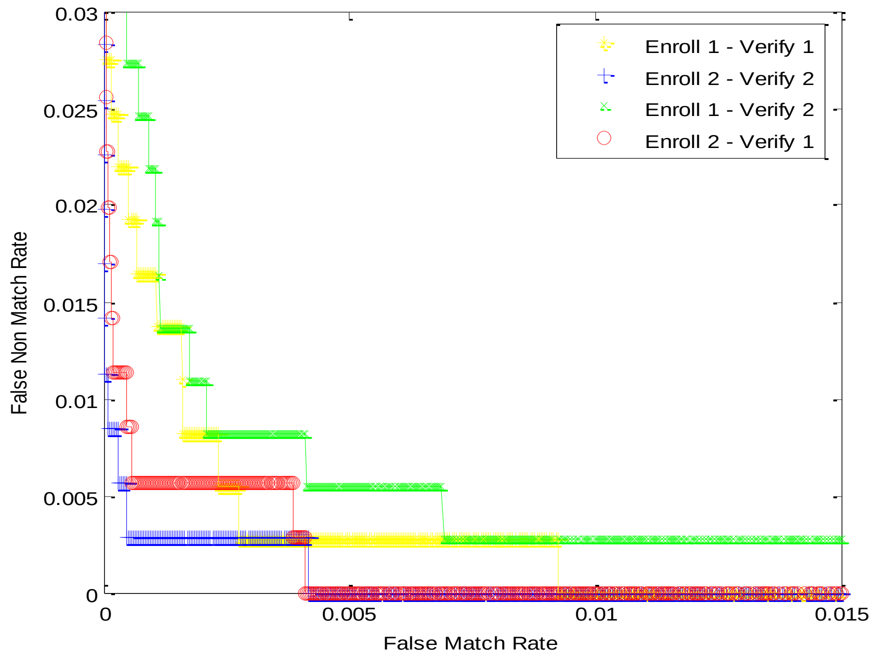


Figure 47. Enroll C Verify B – GFRR Versus GFAR

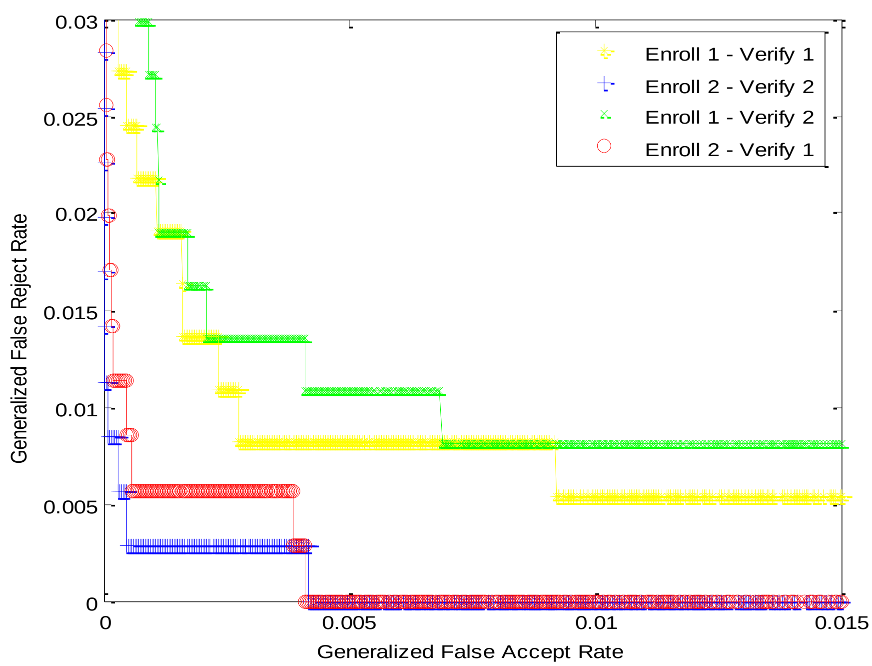


Figure 48. Enroll C Verify C – FNMR Versus FMR

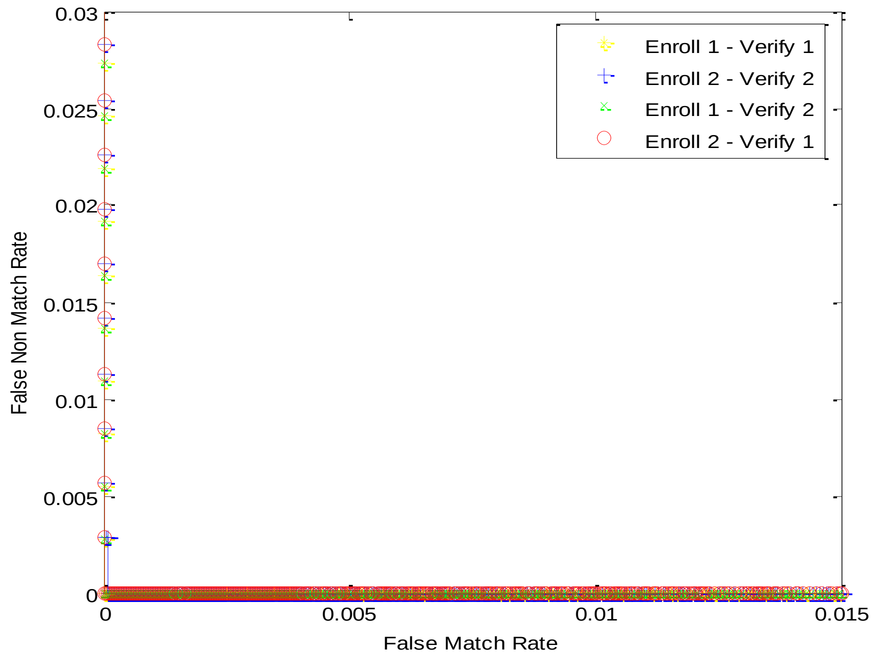


Figure 49. Enroll C Verify C – GFRR Versus GFAR

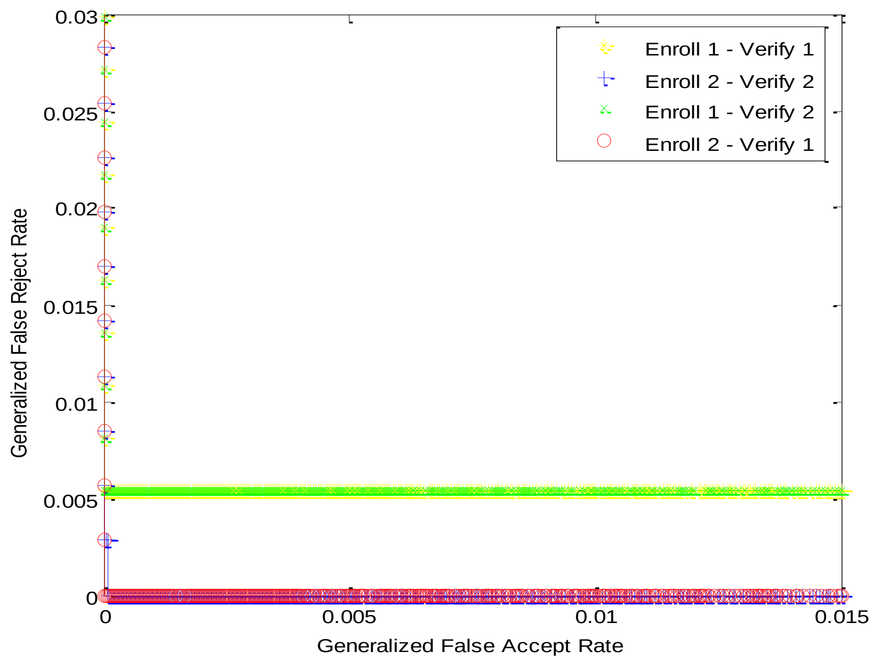


Figure 50. Enroll C Verify D – FNMR Versus FMR

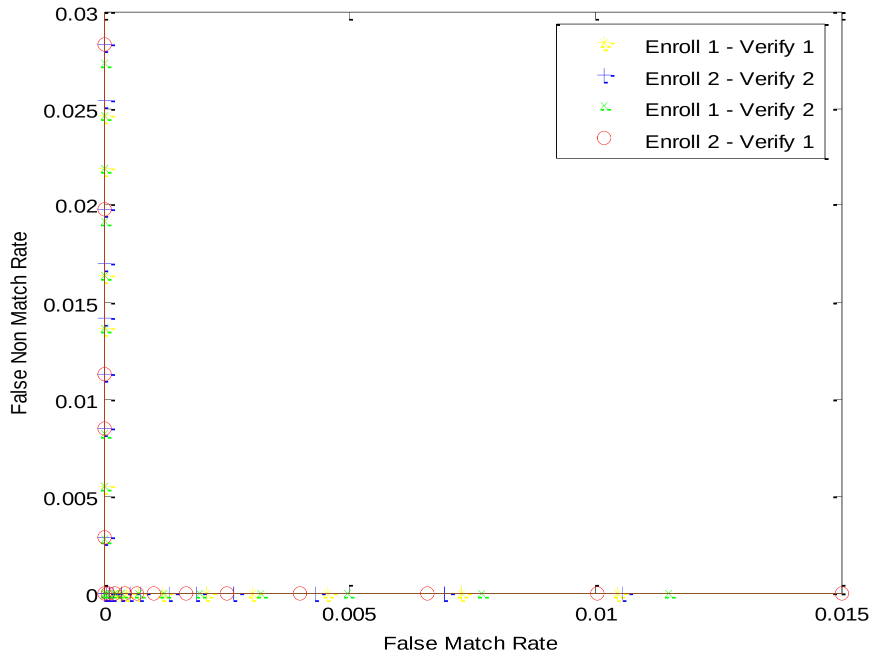


Figure 51. Enroll C Verify D – GFRR Versus GFAR

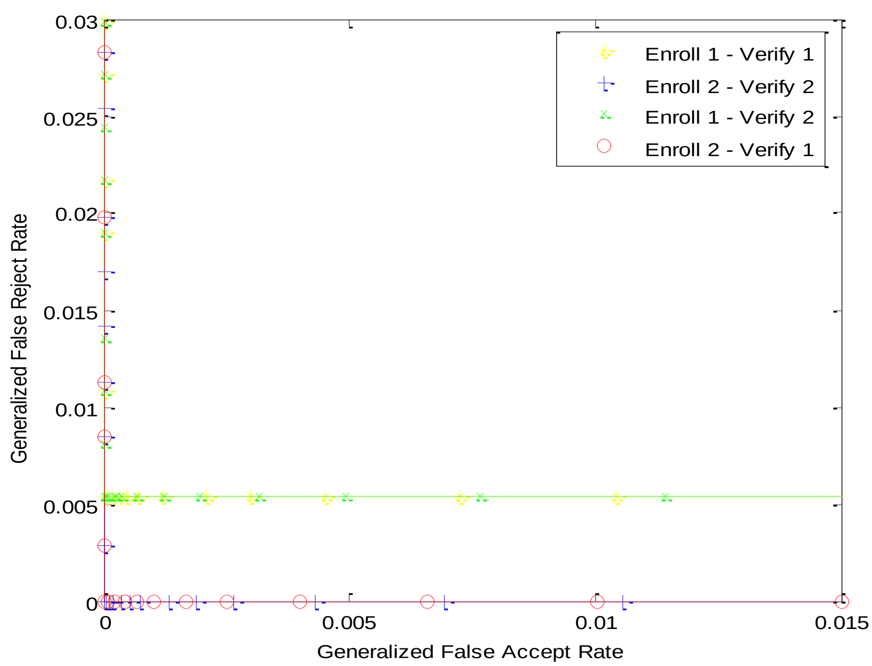


Figure 52. Enroll C Verify E – FNMR Versus FMR

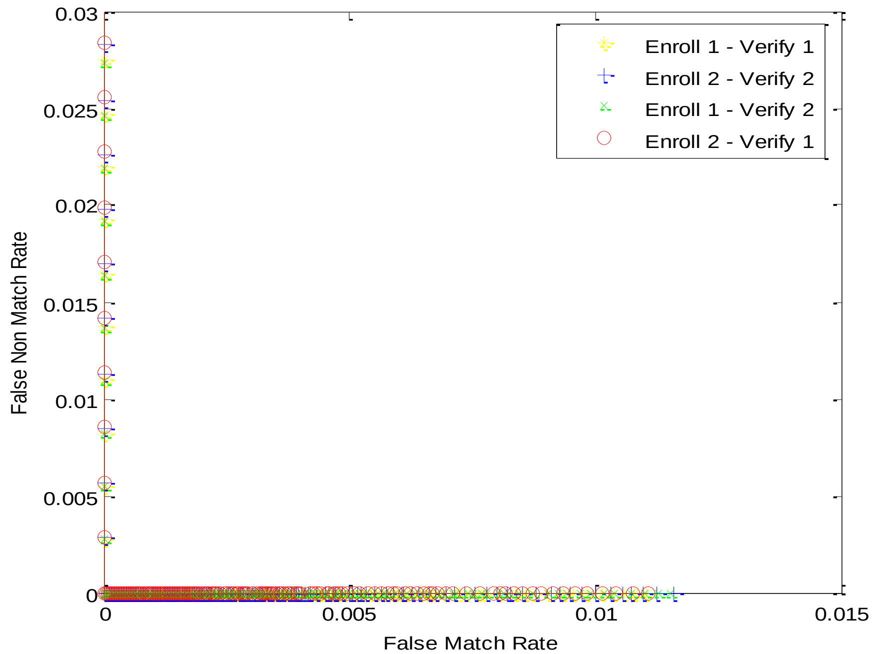


Figure 53. Enroll C Verify E – GFRR Versus GFAR

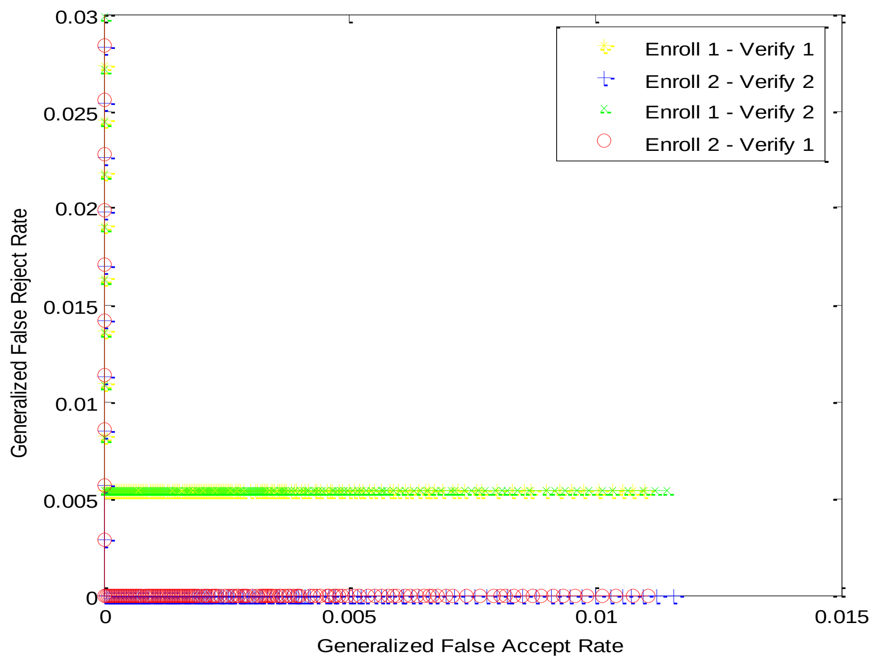


Figure 54. Enroll C Verify F – FNMR Versus FMR

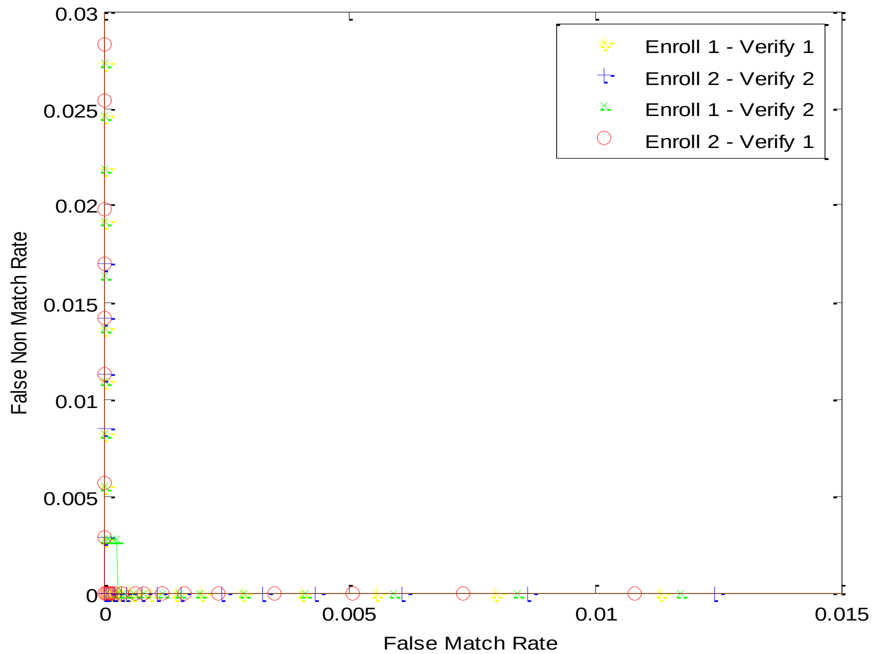


Figure 55. Enroll C Verify F – GFRR Versus GFAR

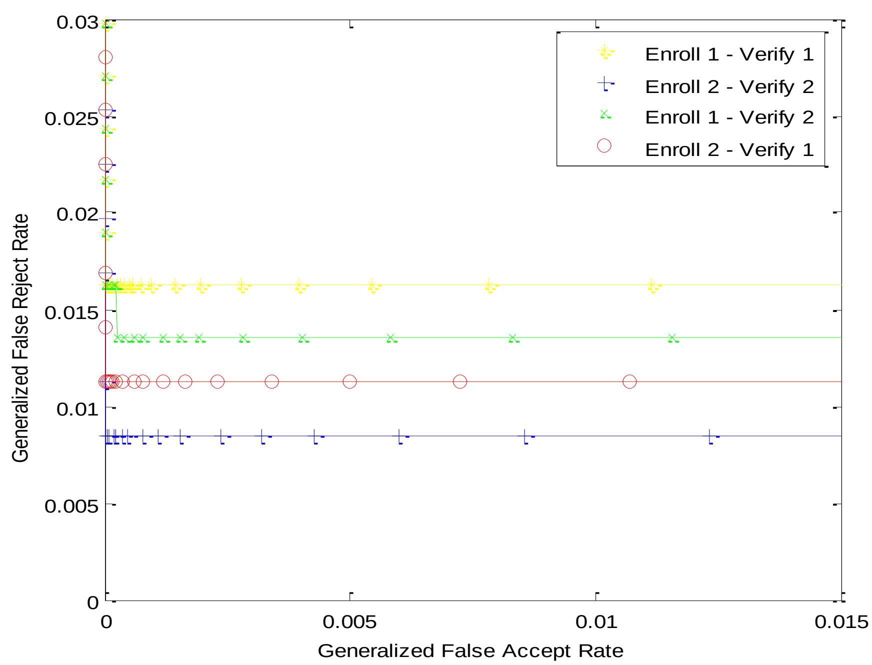


Figure 56. Enroll C Verify G – FNMR Versus FMR

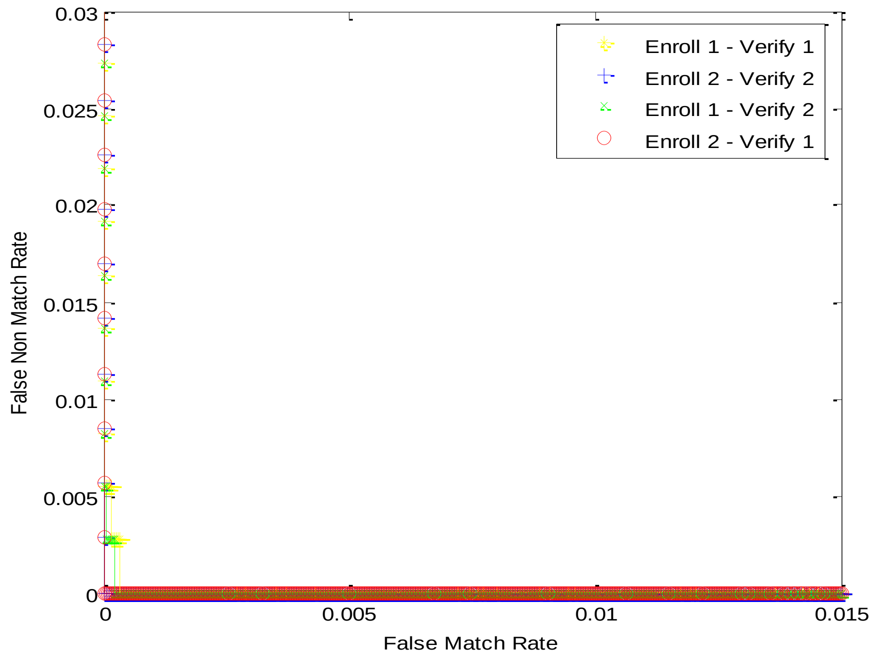


Figure 57. Enroll C Verify G – GFRR Versus GFAR

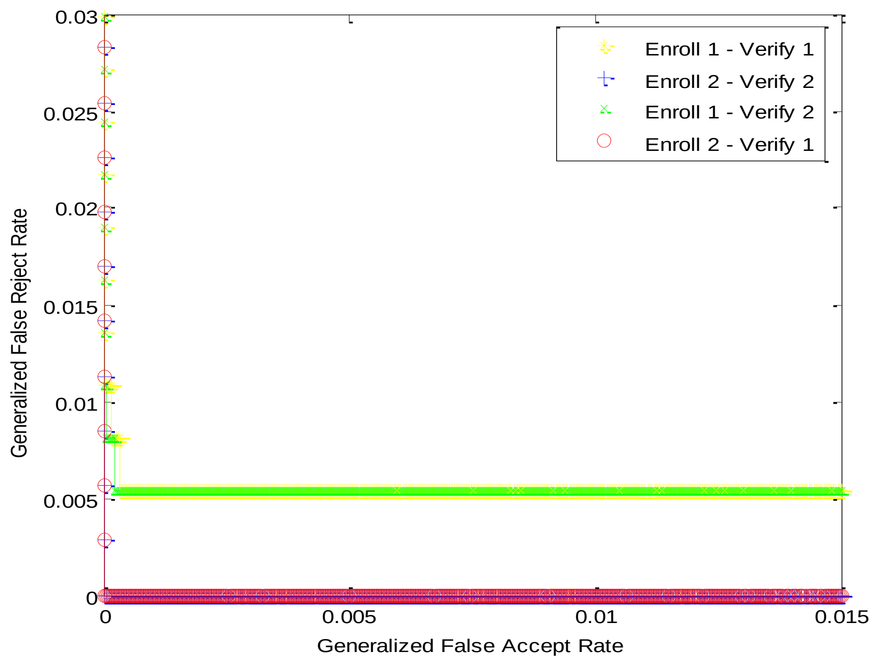


Figure 58. Enroll C Verify H – FNMR Versus FMR

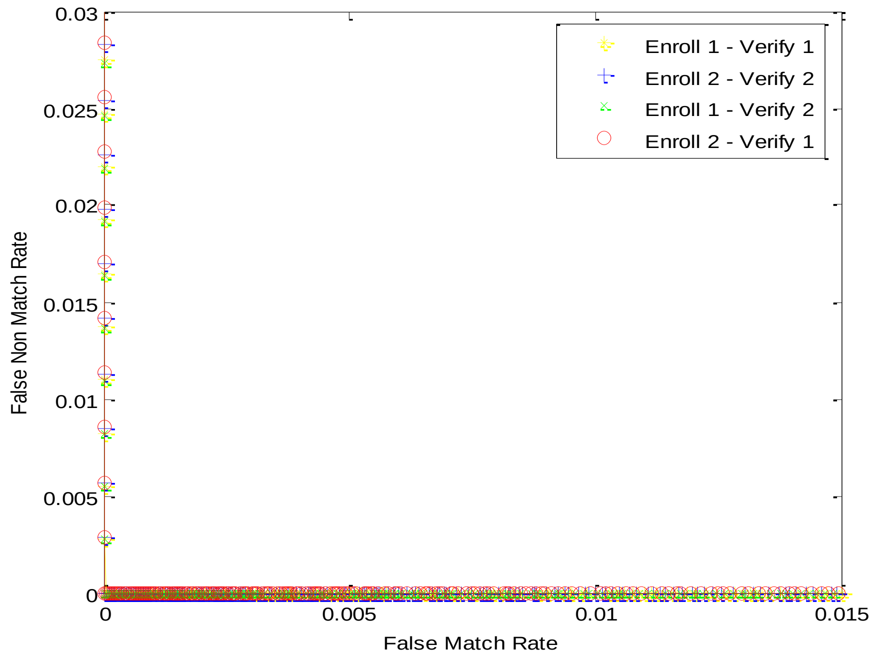


Figure 59. Enroll C Verify H– GFRR Versus GFAR

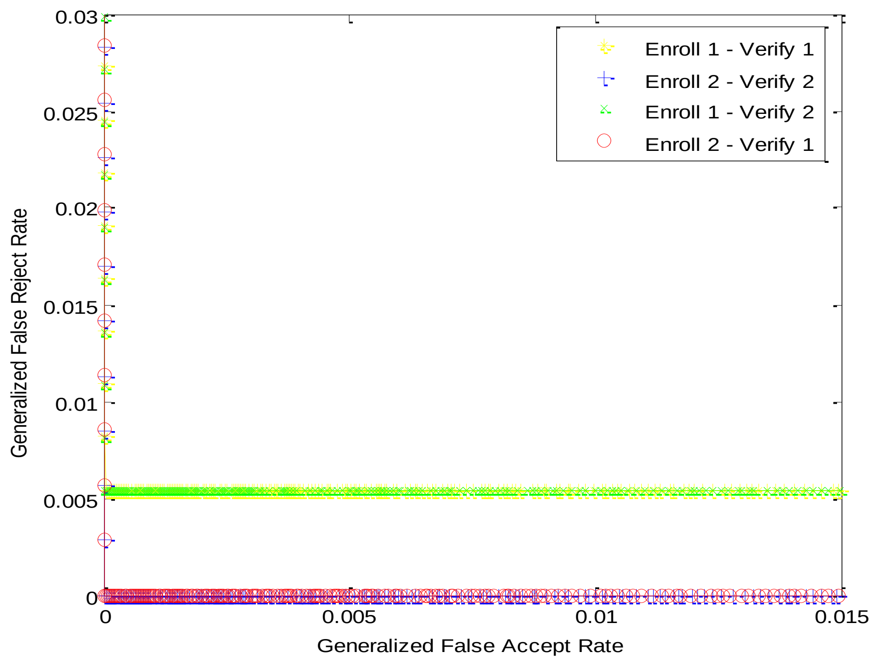


Figure 60. Enroll C Verify I – FNMR Versus FMR

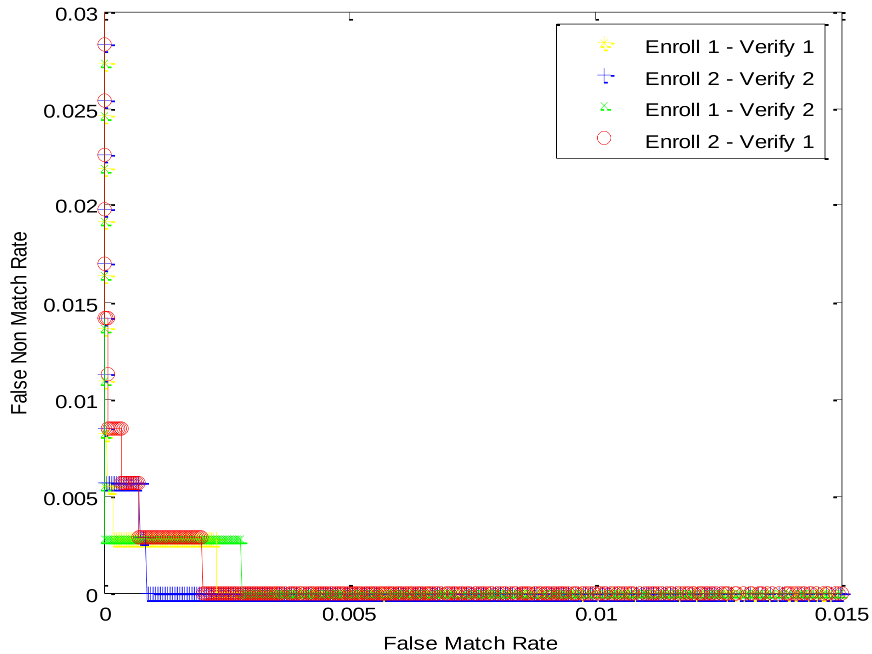


Figure 61. Enroll C Verify I – GFRR Versus GFAR

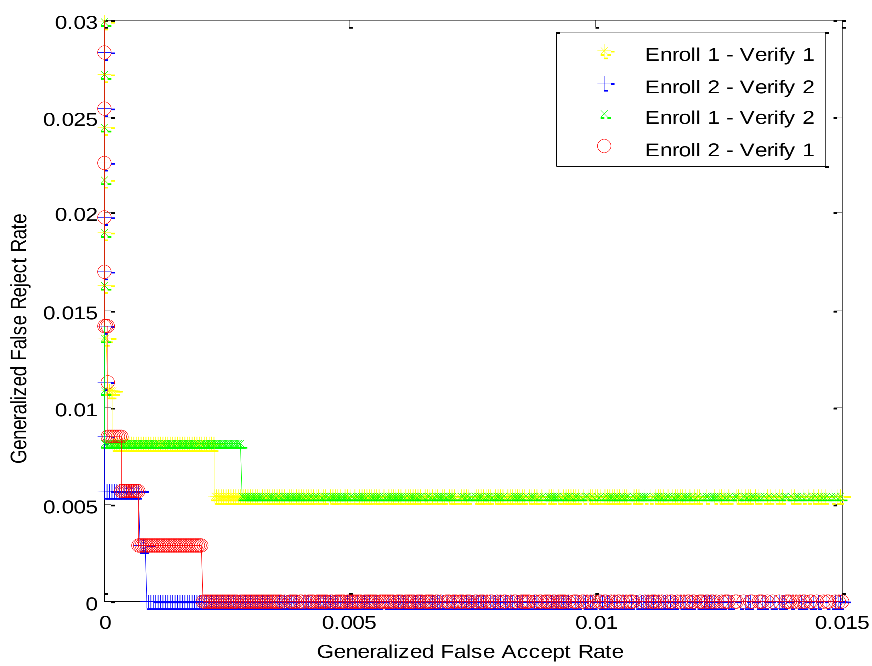


Figure 62. Enroll D Verify A1– FNMR Versus FMR

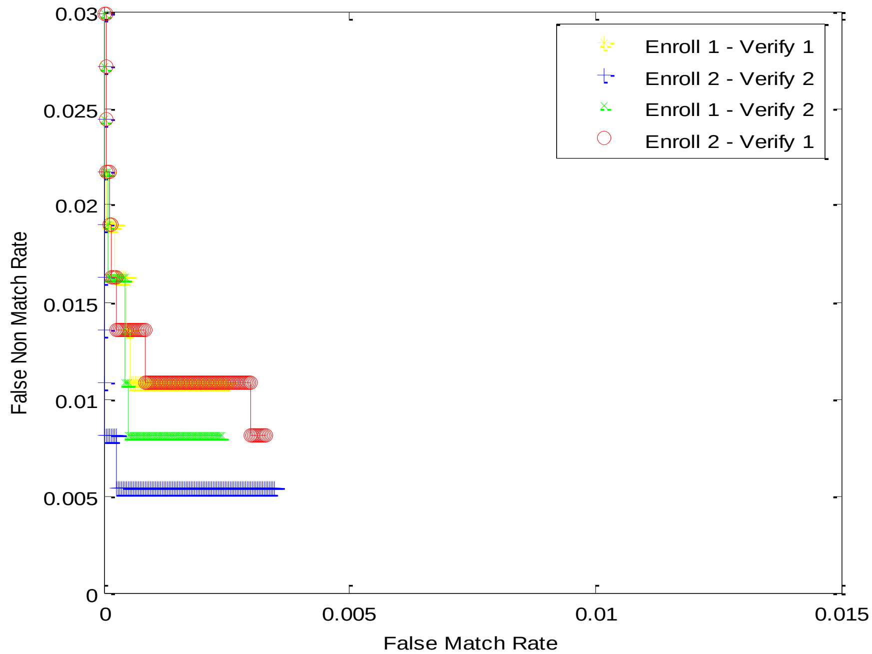


Figure 63. Enroll D Verify A1 – GFRR Versus GFAR

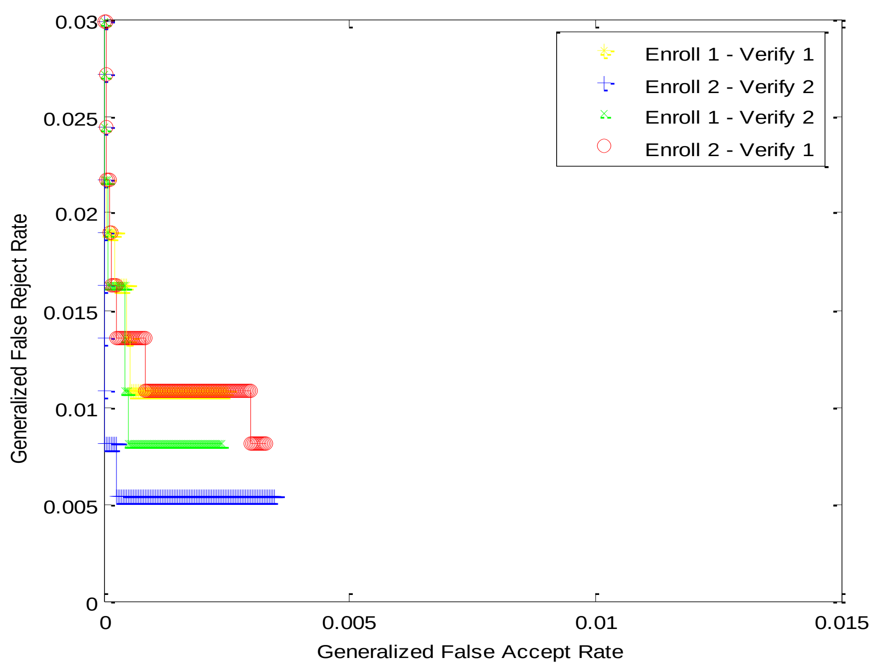


Figure 64. Enroll D Verify B – FNMR Versus FMR

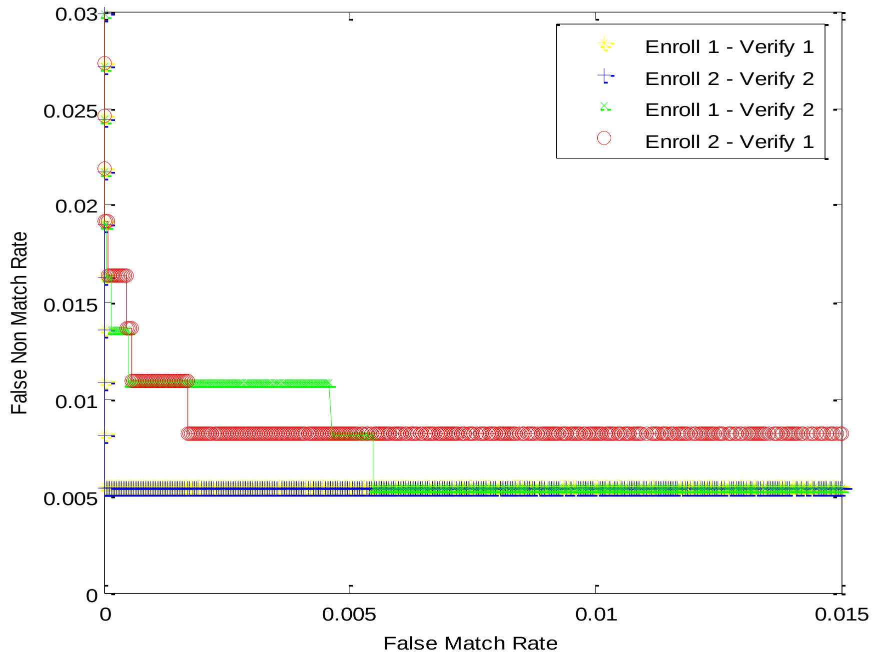


Figure 65. Enroll D Verify B – GFRR Versus GFAR

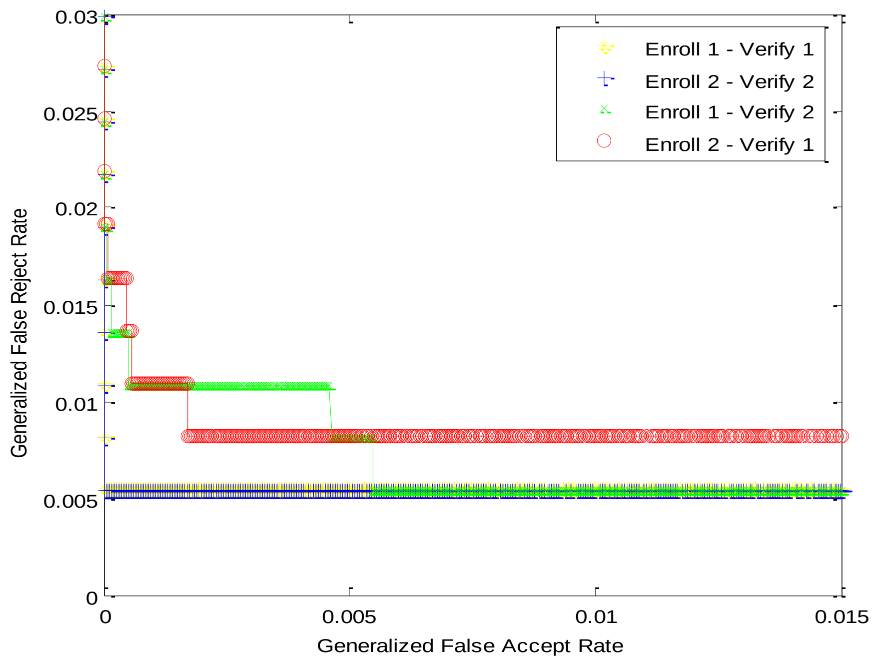


Figure 66. Enroll D Verify C – FNMR Versus FMR

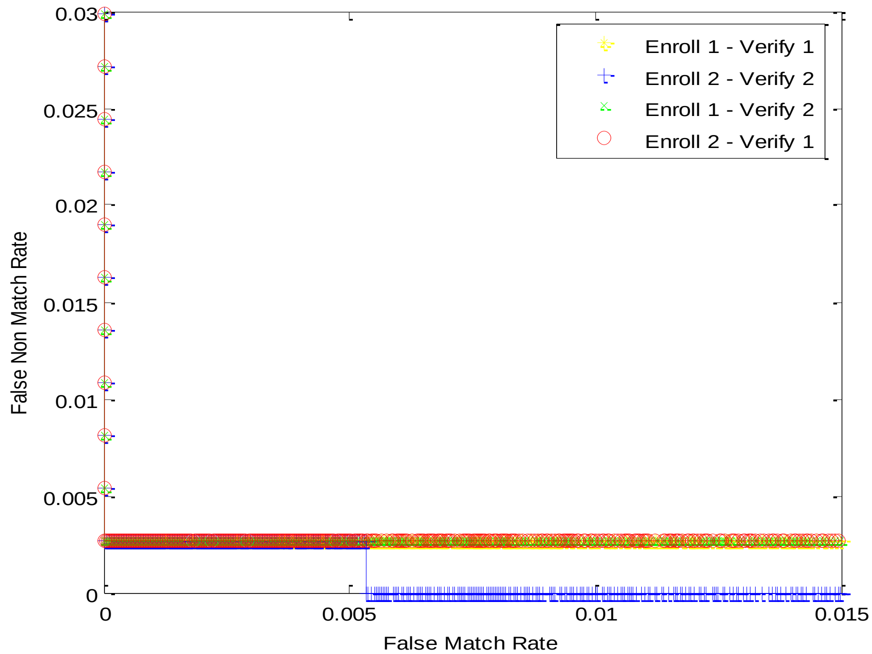


Figure 67. Enroll D Verify C – GFRR Versus GFAR

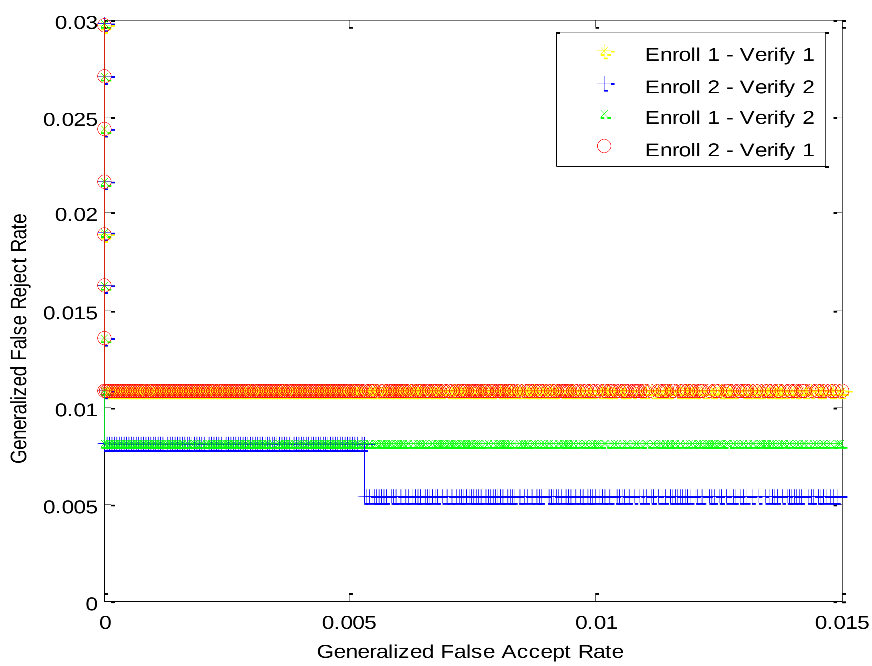


Figure 68. Enroll D Verify D – FNMR Versus FMR

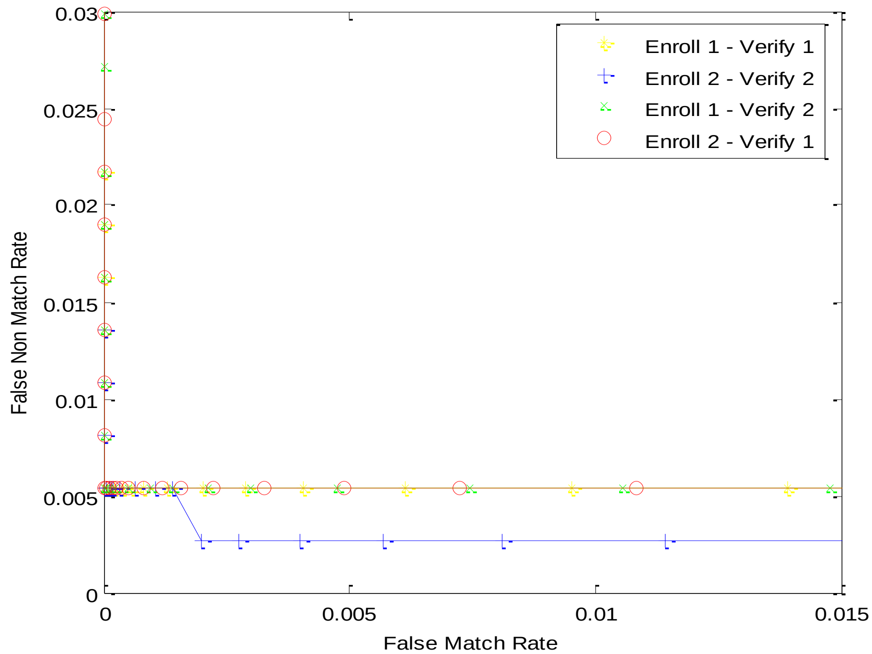


Figure 69. Enroll D Verify D – GFRR Versus GFAR

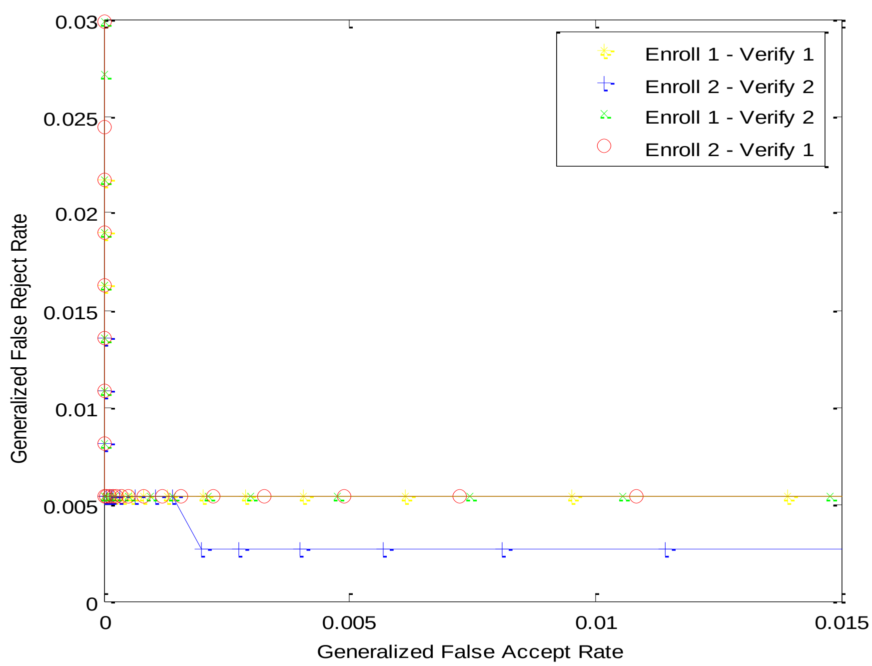


Figure 70. Enroll D Verify E – FNMR Versus FMR

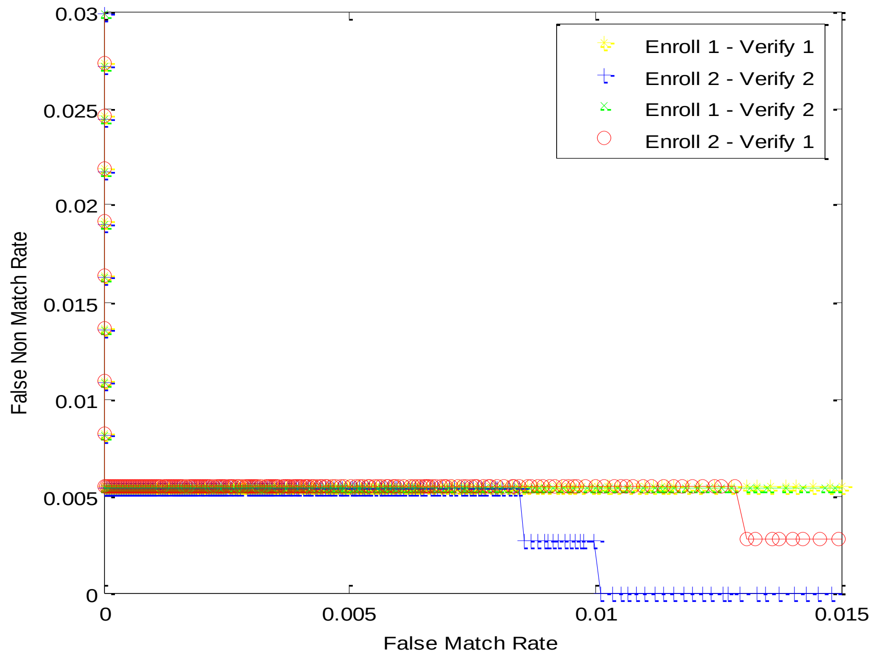


Figure 71. Enroll D Verify E – GFRR Versus GFAR

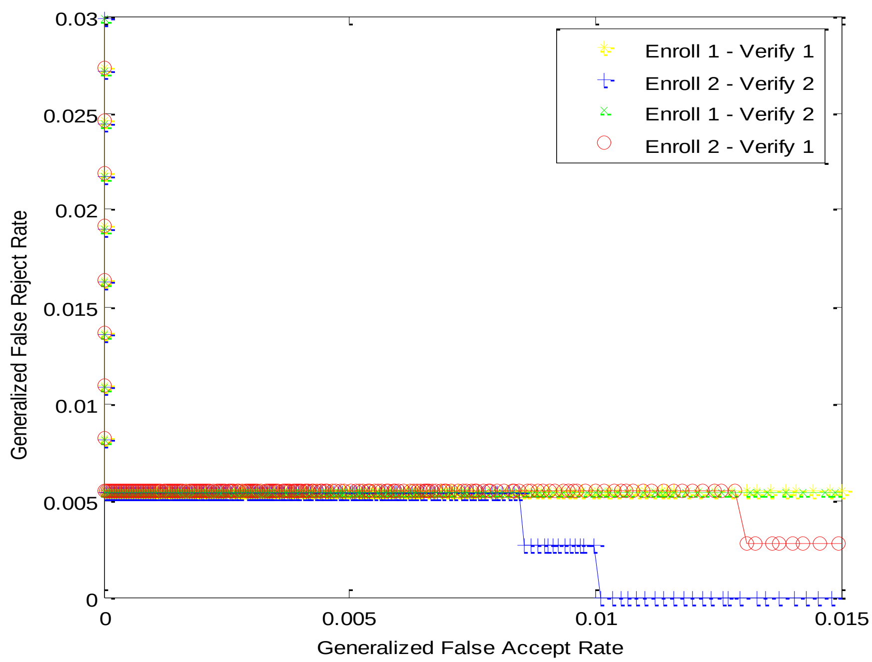


Figure 72. Enroll D Verify F – FNMR Versus FMR

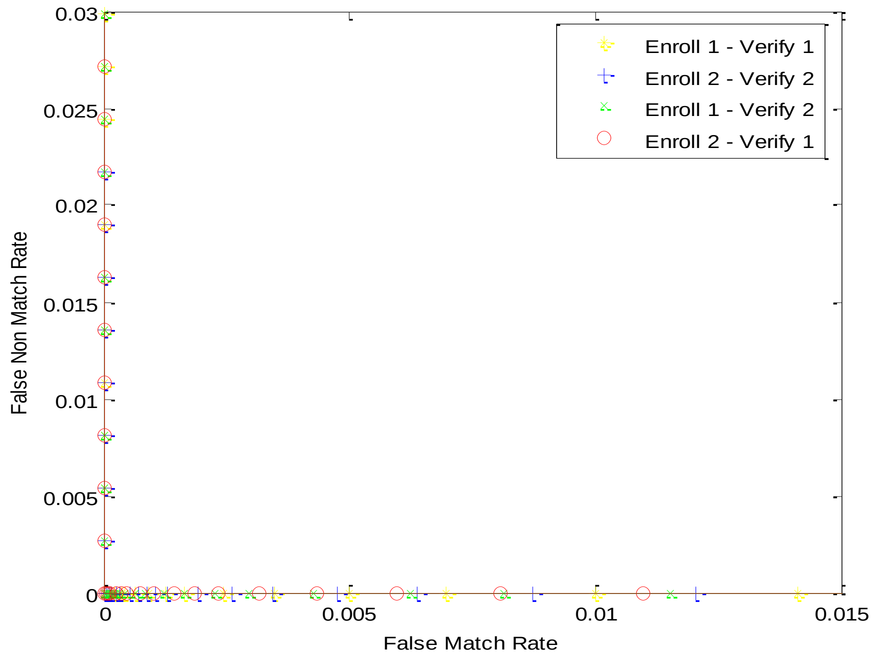


Figure 73. Enroll D Verify F – GFRR Versus GFAR

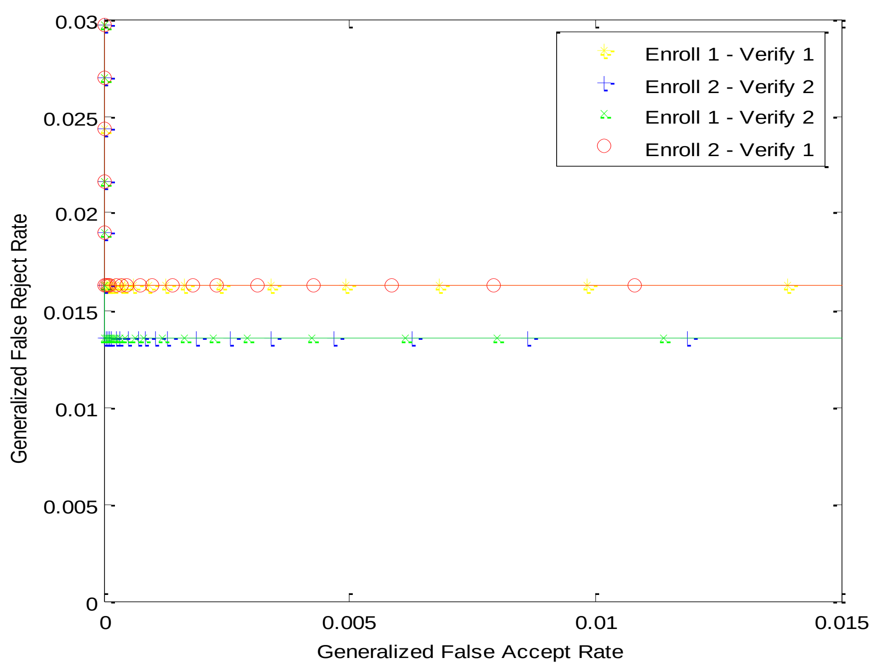


Figure 74. Enroll D Verify G – FNMR Versus FMR

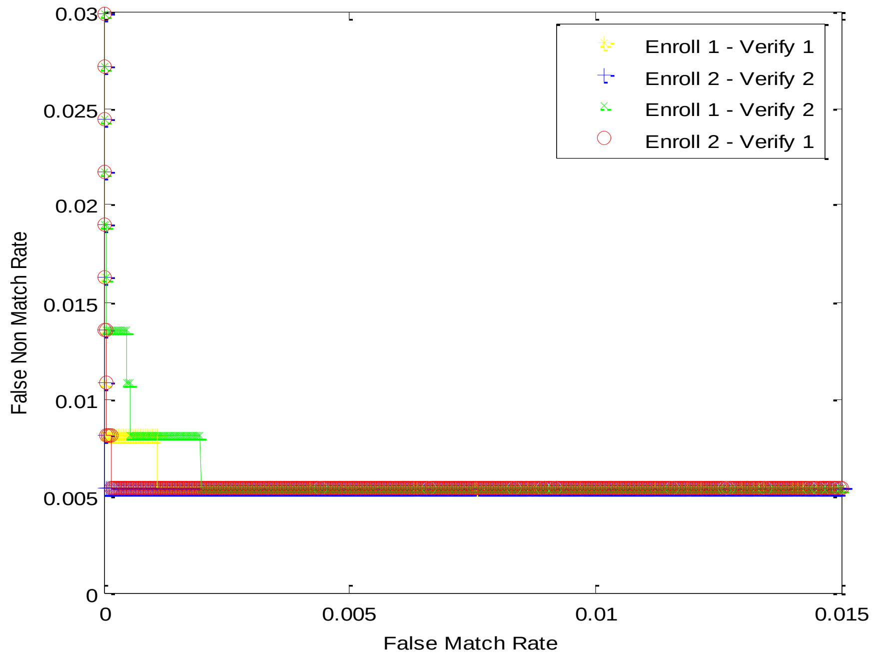


Figure 75. Enroll D Verify G – GFRR Versus GFAR

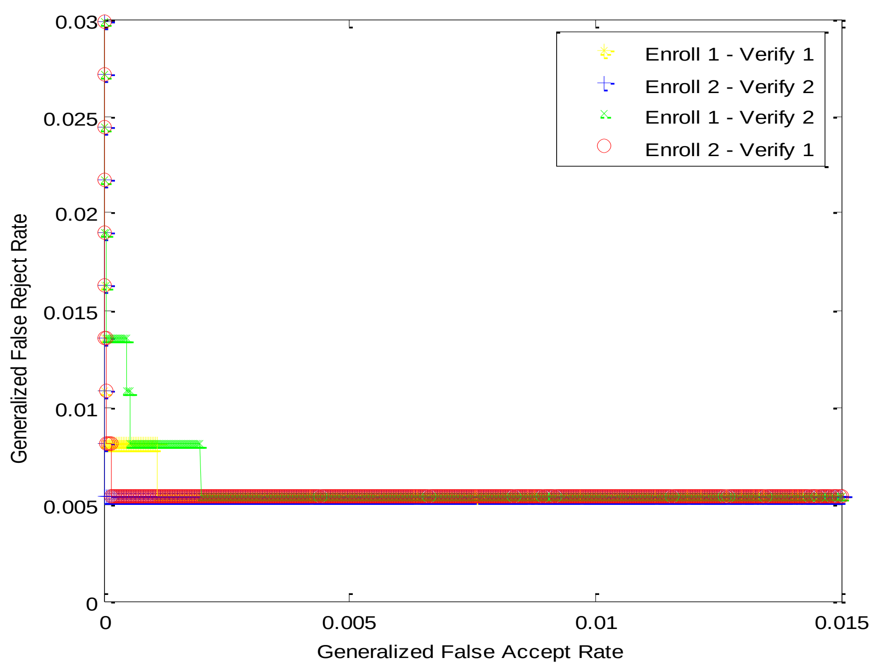


Figure 76. Enroll D Verify H – FNMR Versus FMR

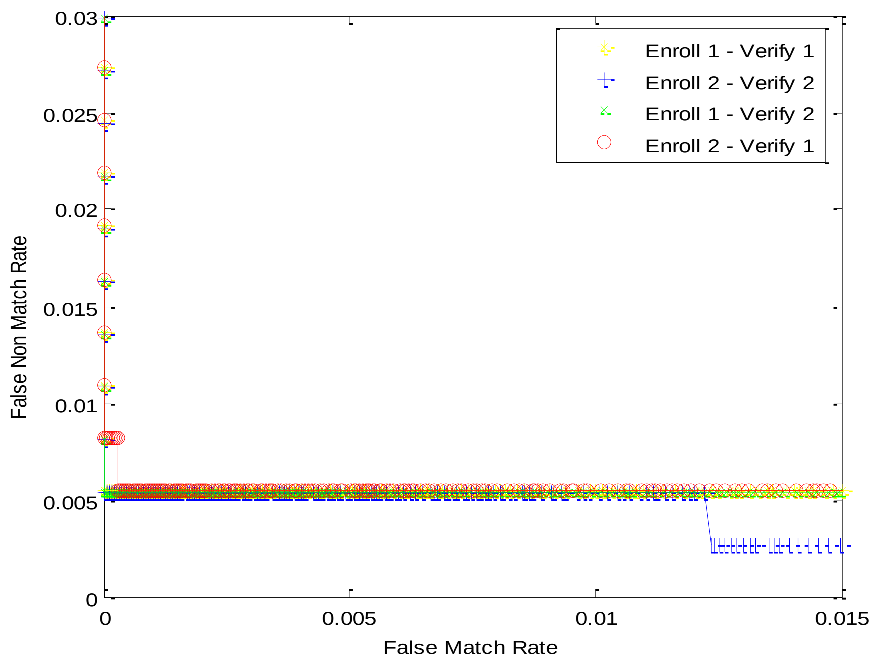


Figure 77. Enroll D Verify H – GFRR Versus GFAR

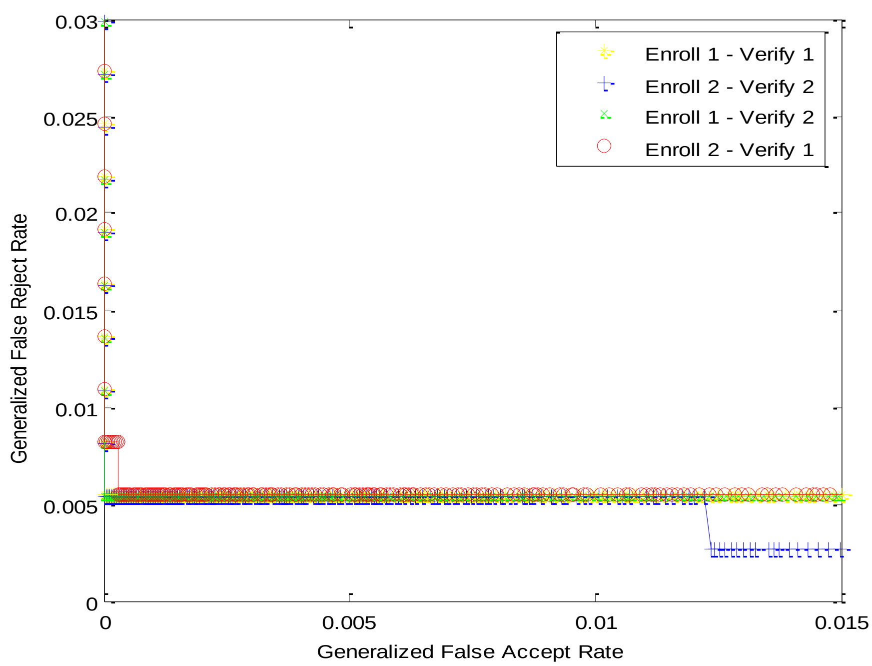


Figure 78. Enroll D Verify I – FNMR Versus FMR

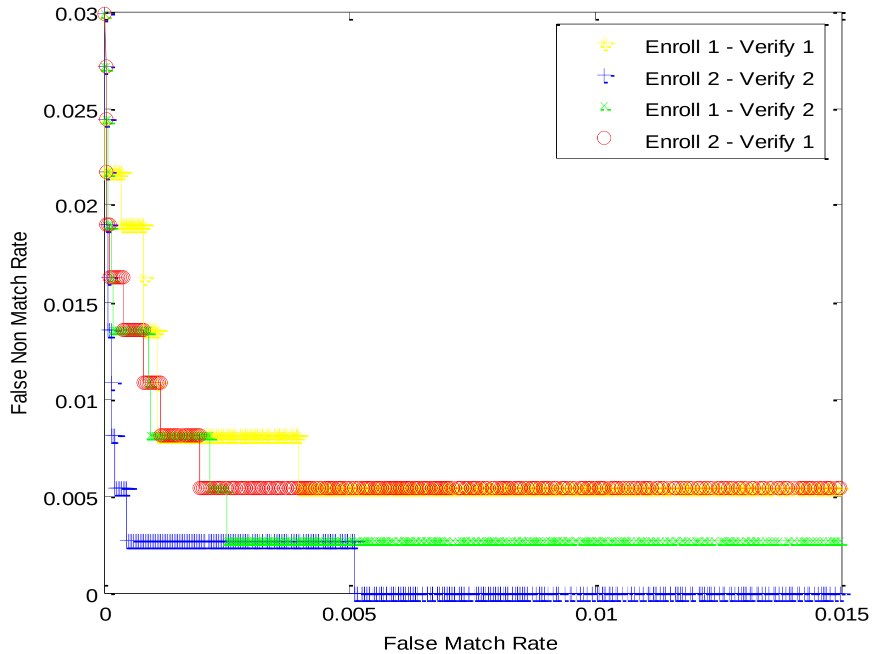


Figure 79. Enroll D Verify I – GFRR Versus GFAR

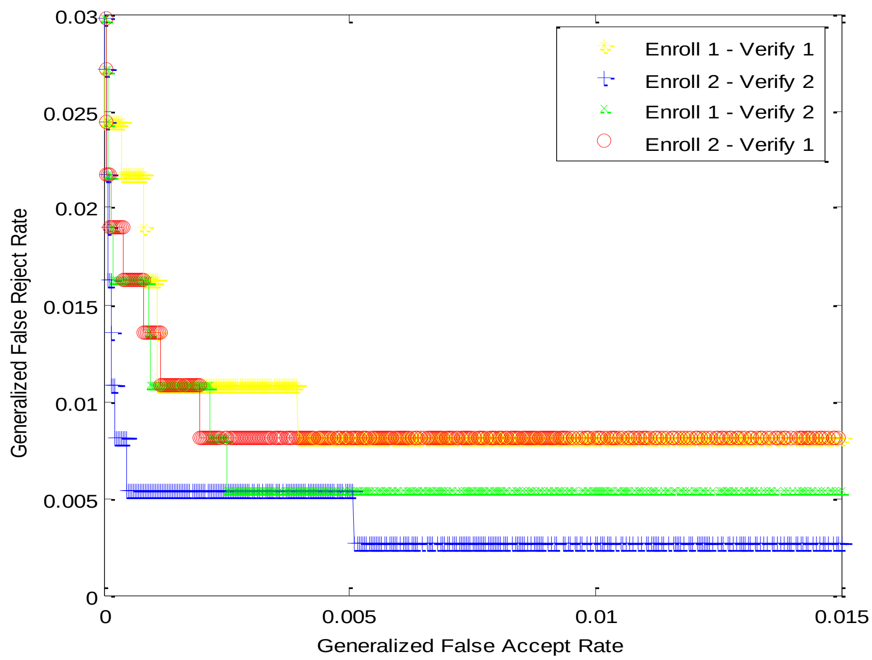


Figure 80. Enroll E Verify A1 – FNMR Versus FMR

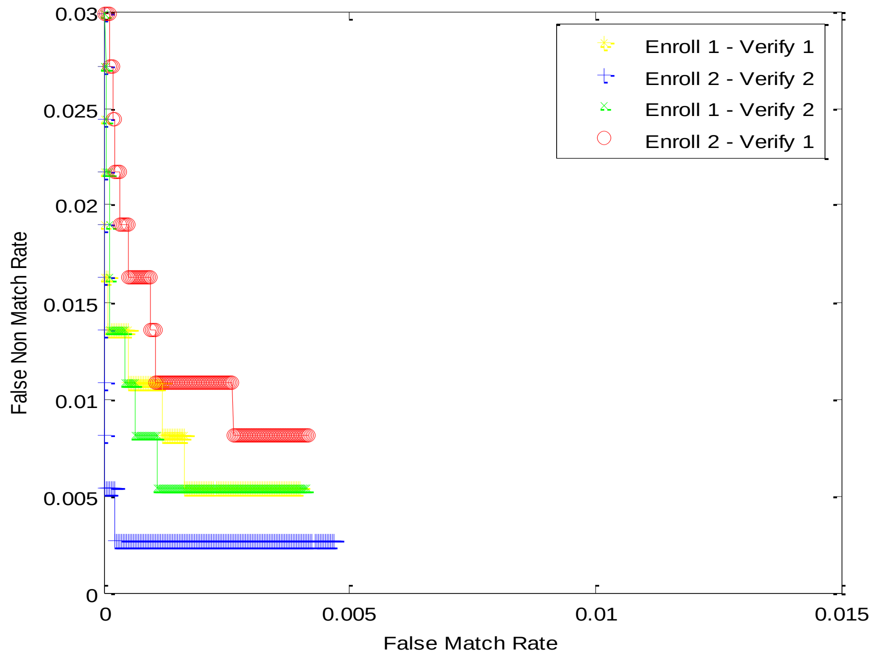


Figure 81. Enroll E Verify A1 – GFRR Versus GFAR

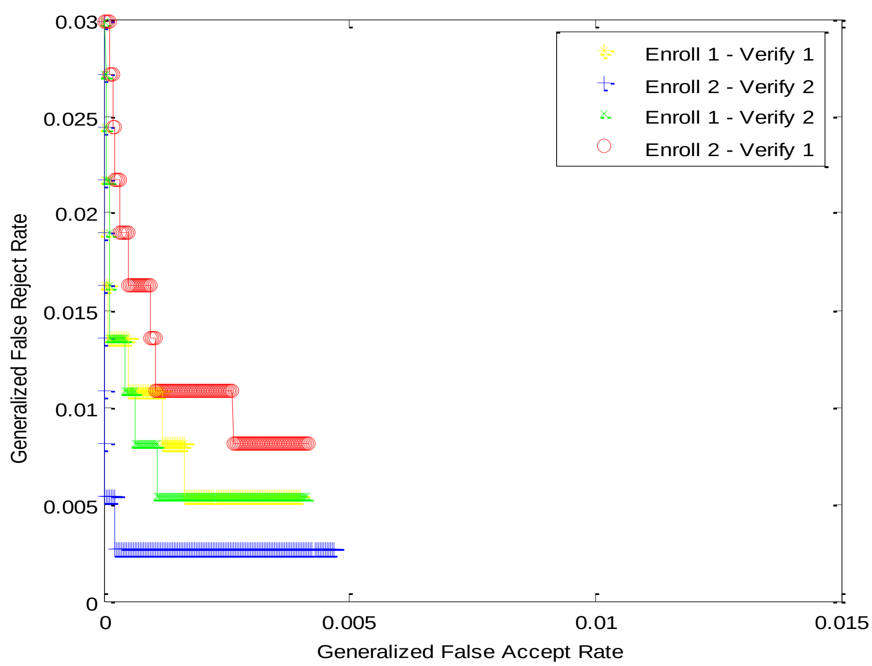


Figure 82. Enroll E Verify B – FNMR Versus FMR

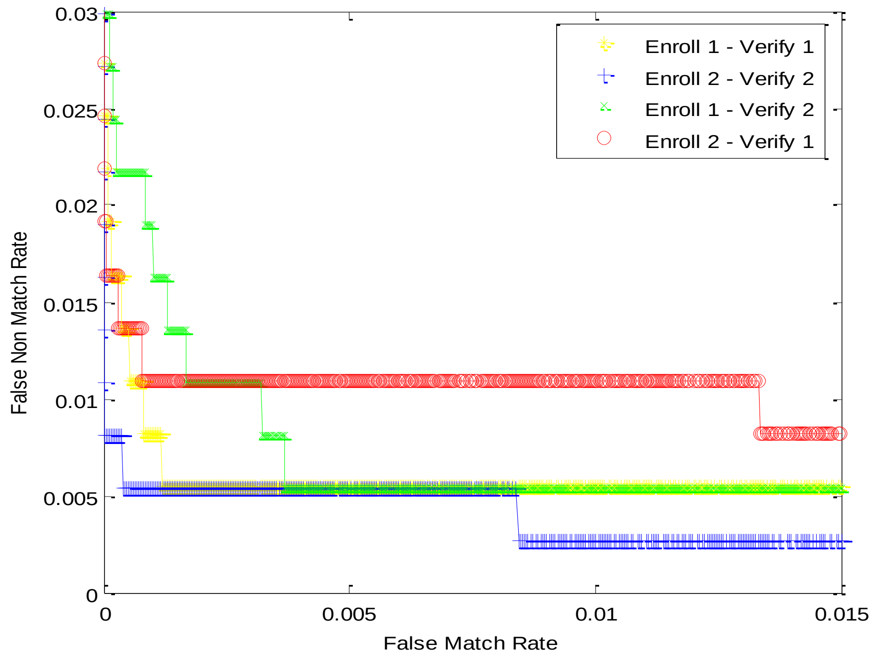


Figure 83. Enroll E Verify B – GFRR Versus GFAR

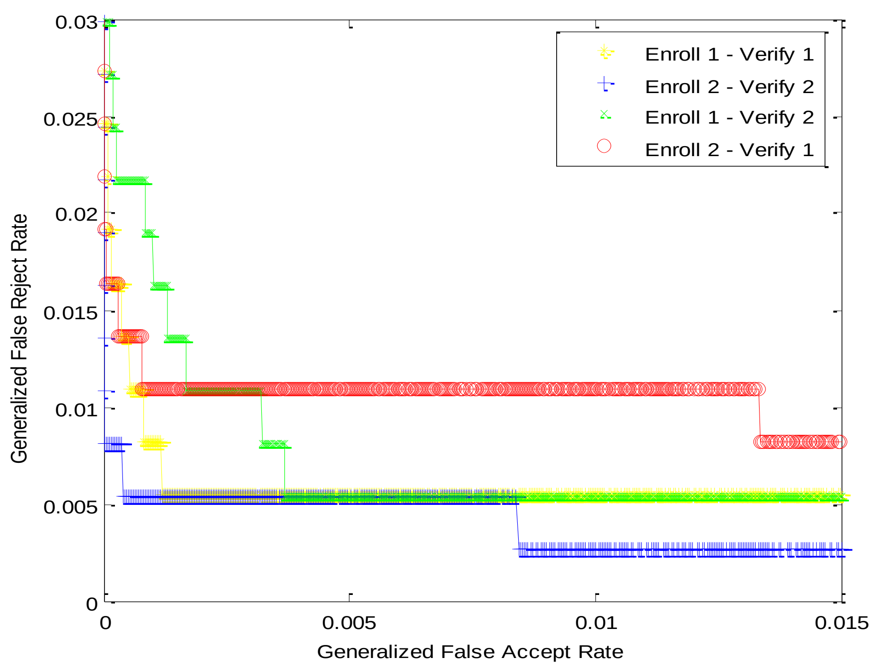


Figure 84. Enroll E Verify C – FNMR Versus FMR

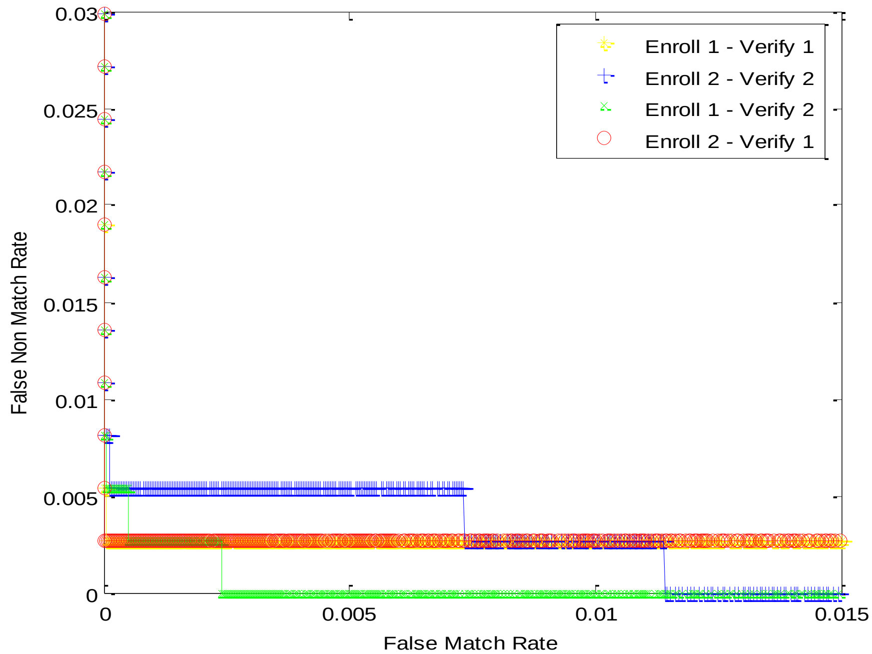


Figure 85. Enroll E Verify C – GFRR Versus GFAR

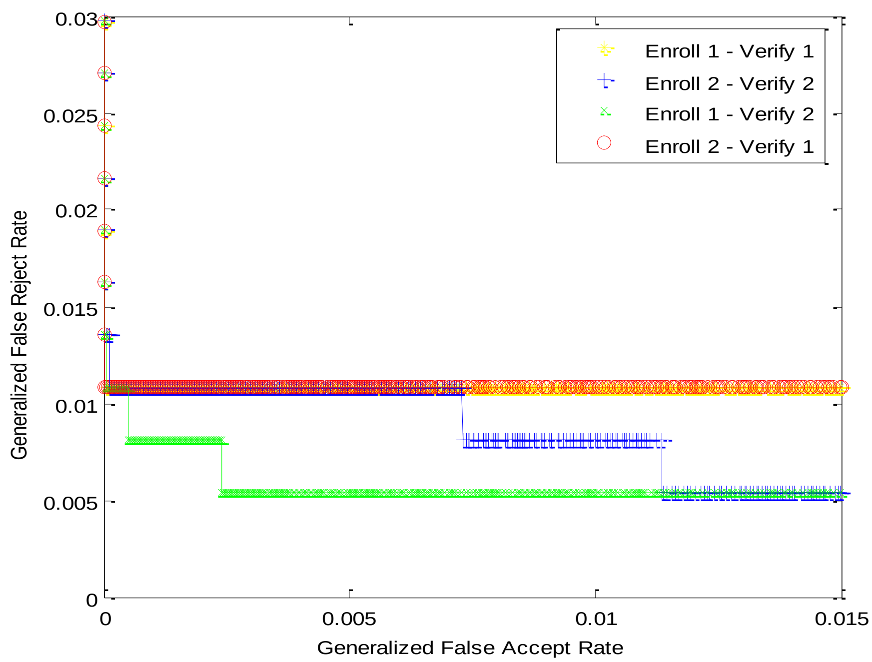


Figure 86. Enroll E Verify D –FNMR Versus FMR

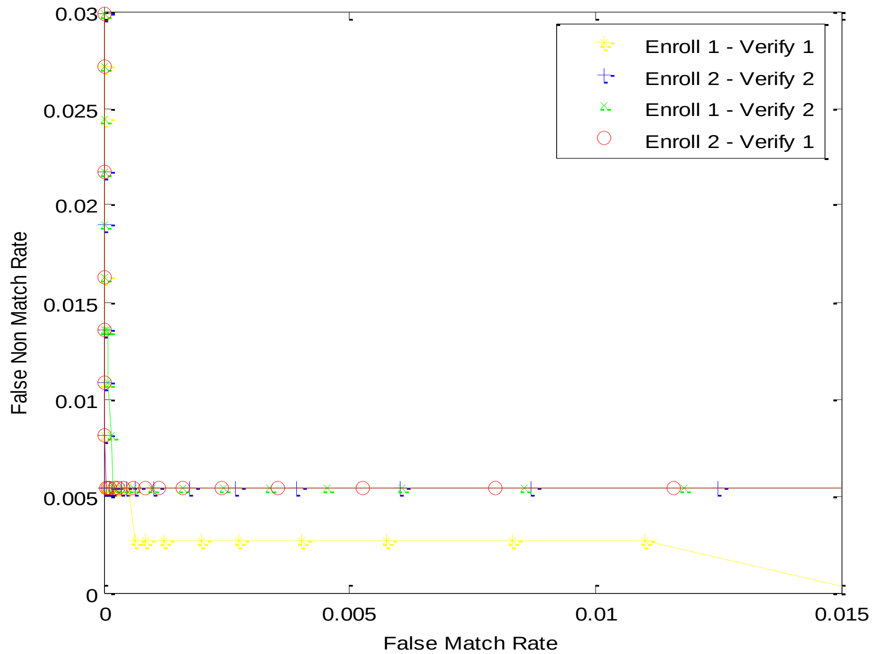


Figure 87. Enroll E Verify D – GFRR Versus GFAR

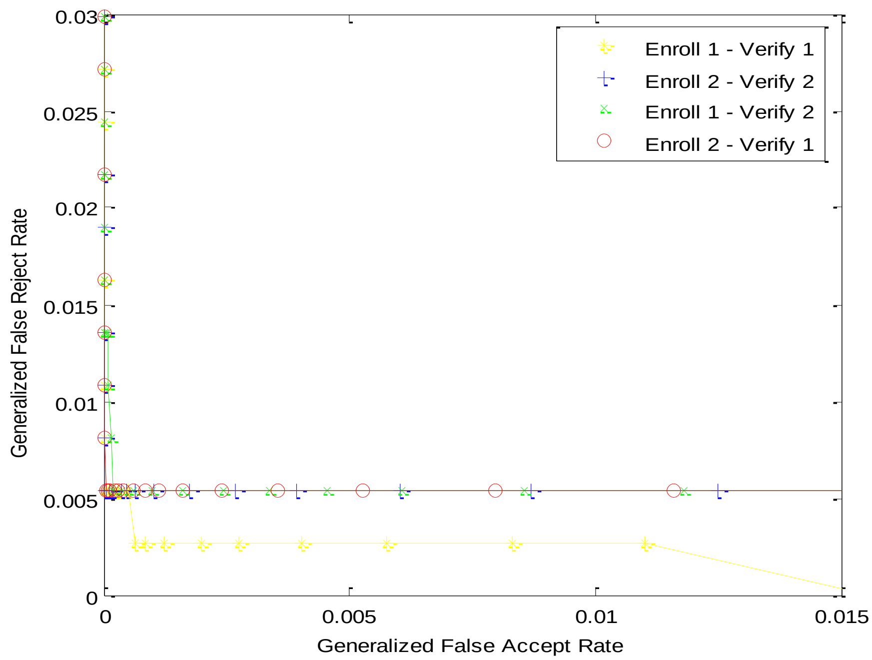


Figure 88. Enroll E Verify E – FNMR Versus FMR

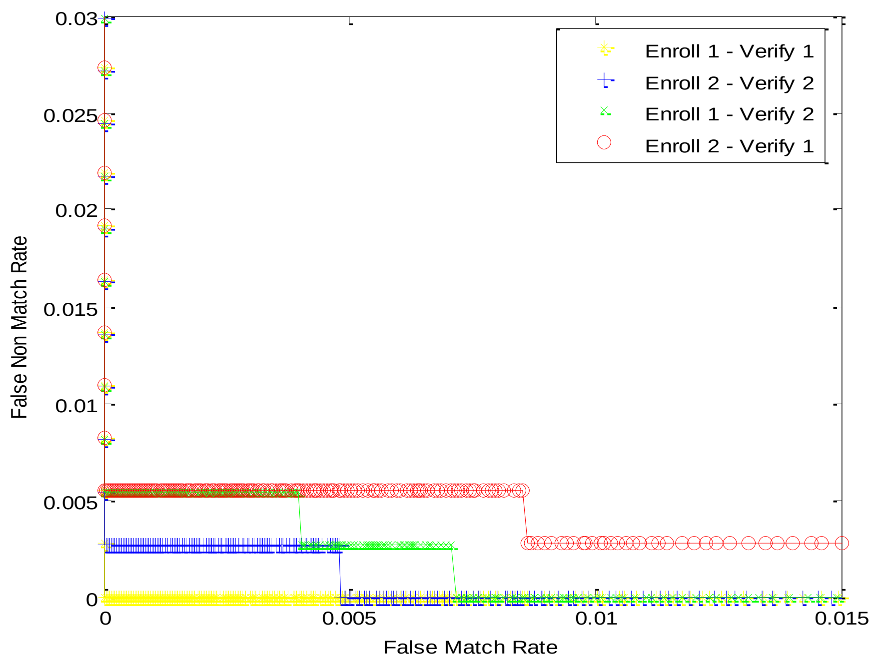


Figure 89. Enroll E Verify E – GFRR Versus GFAR

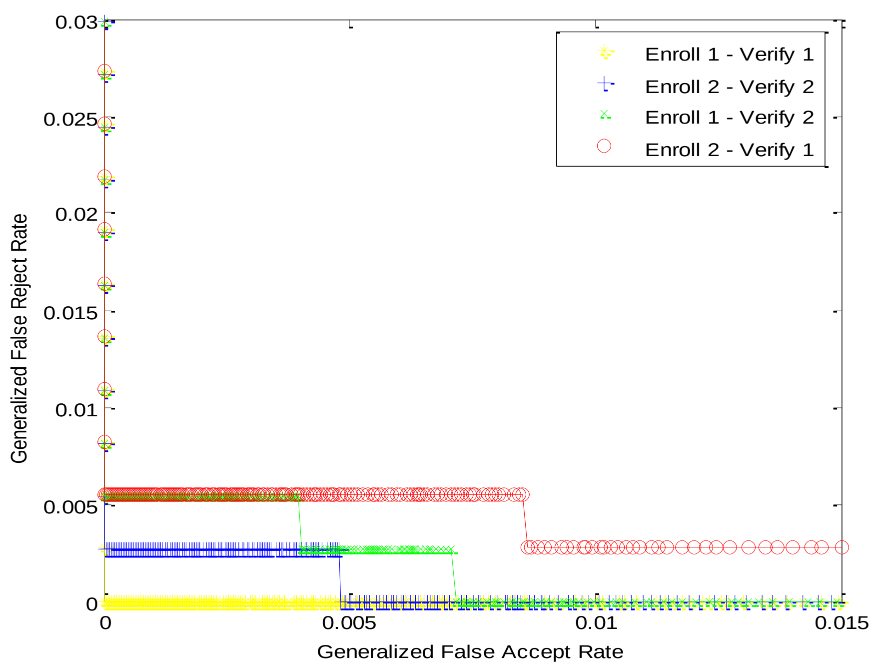


Figure 90. Enroll E Verify F – FNMR Versus FMR

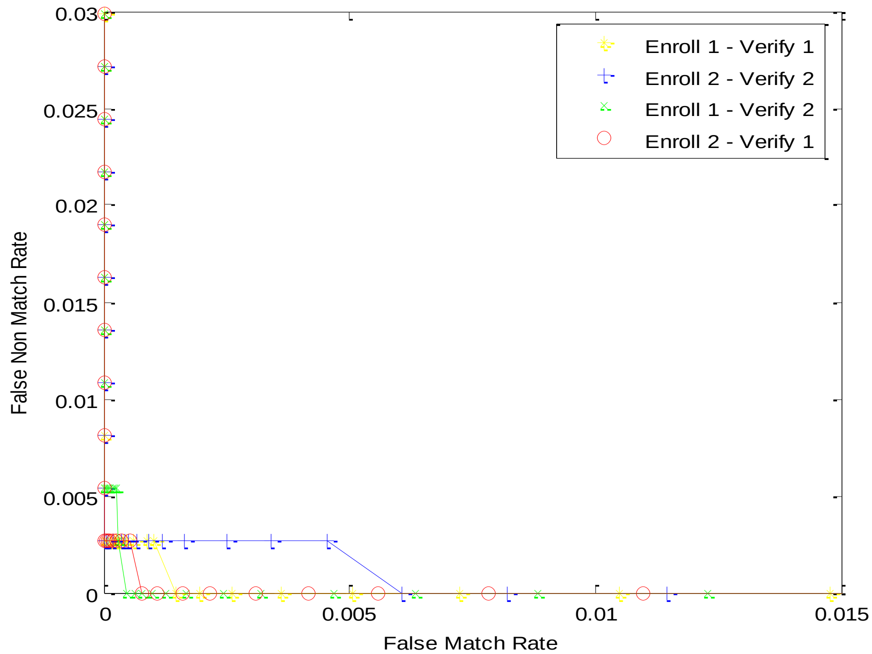


Figure 91. Enroll E Verify F – GFRR Versus GFAR

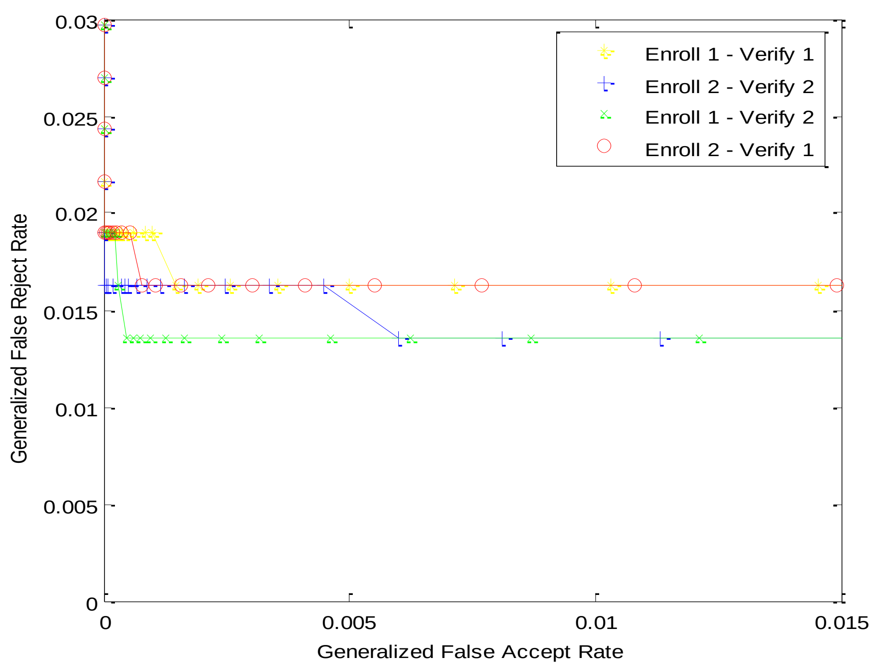


Figure 92. Enroll E Verify G – FNMR Versus FMR

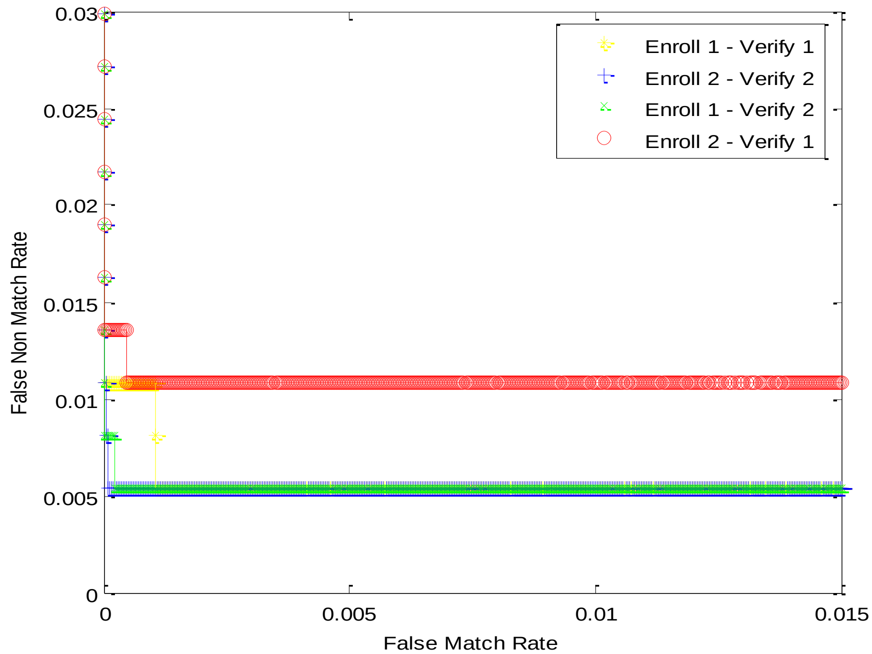


Figure 93. Enroll E Verify G – GFRR Versus GFAR

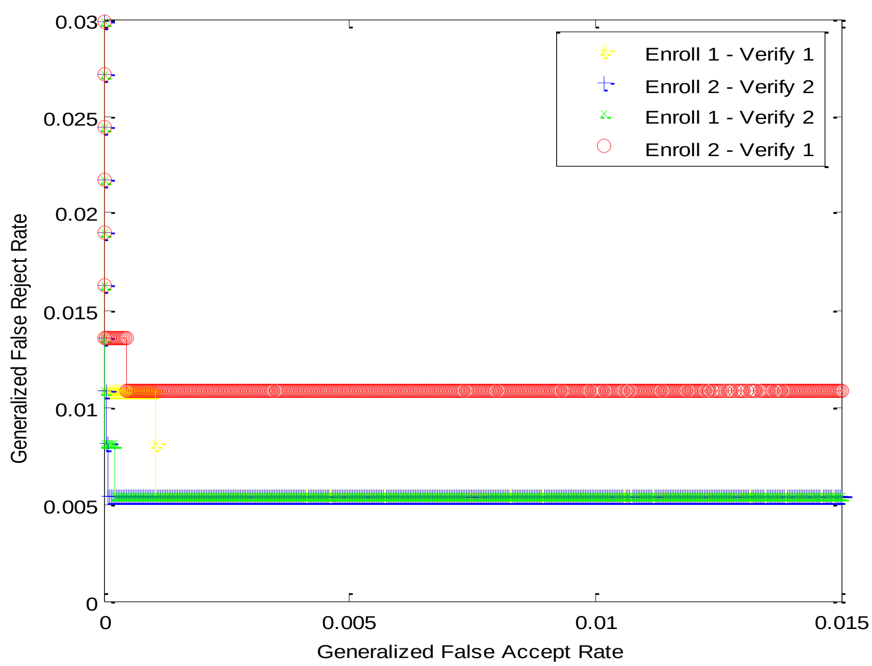


Figure 94. Enroll E Verify H – FNMR Versus FMR

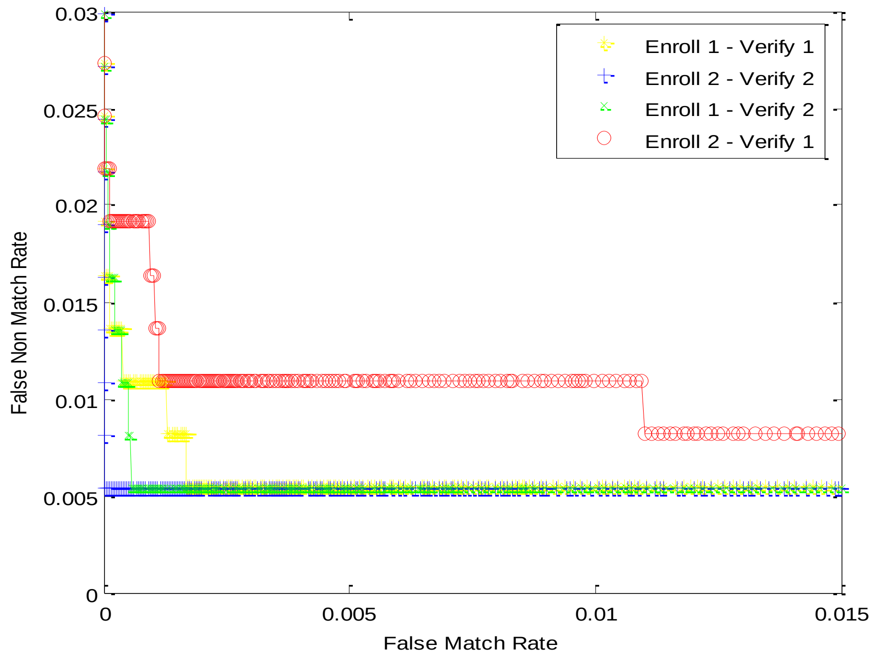


Figure 95. Enroll E Verify H – GFRR Versus GFAR

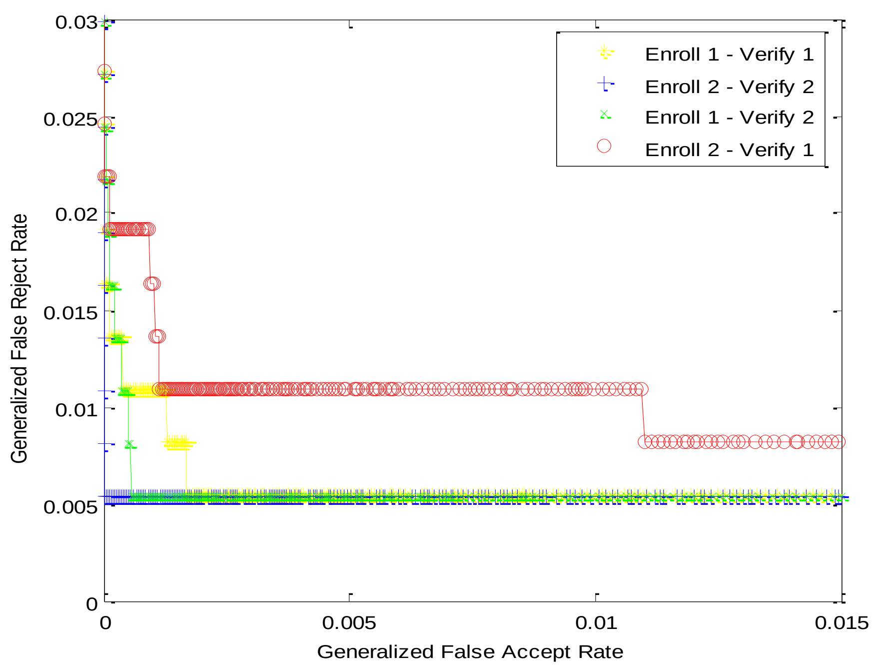


Figure 96. Enroll E Verify I – FNMR Versus FMR

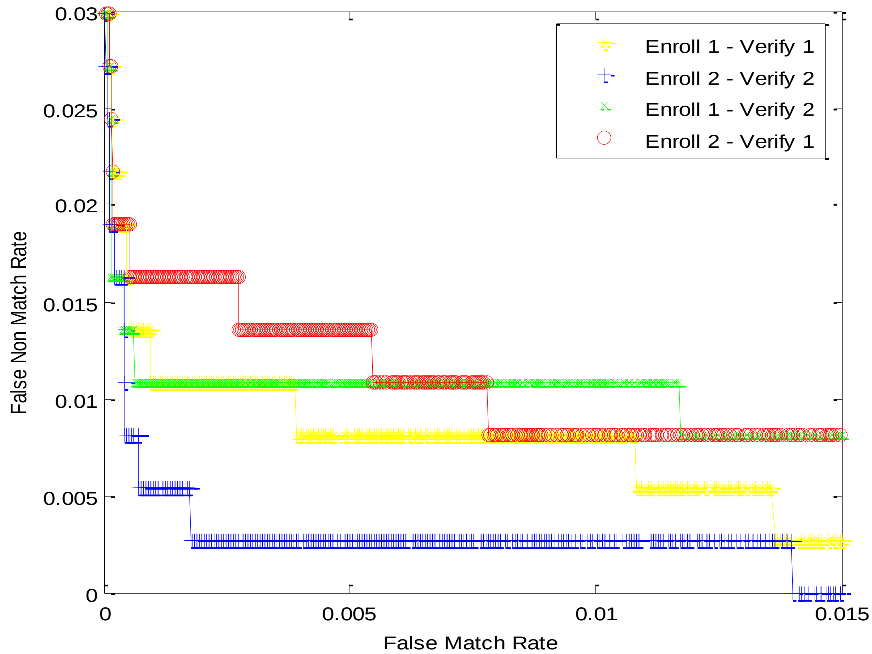


Figure 97. Enroll E Verify I – GFRR Versus GFAR

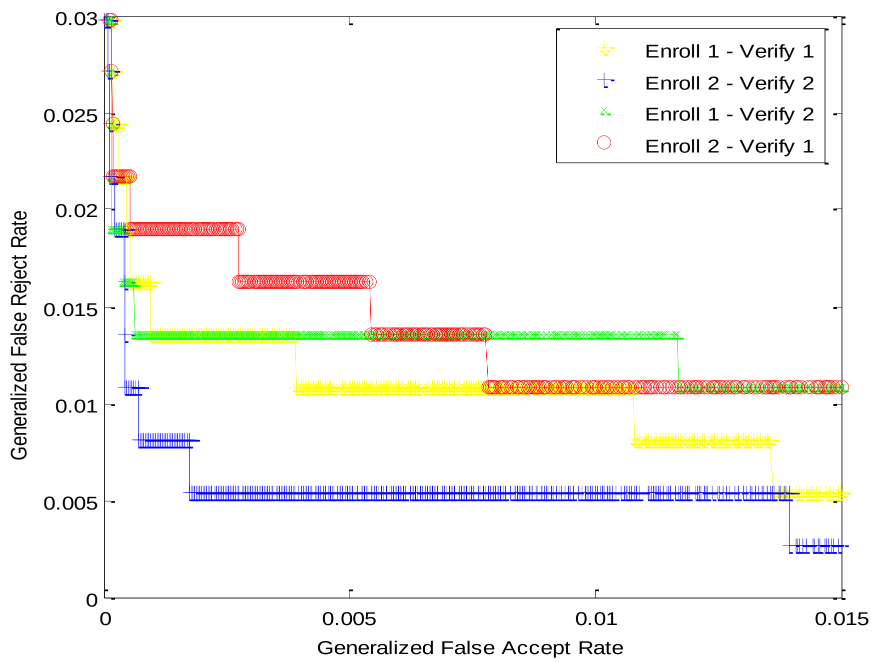


Figure 98. Enroll F Verify A1 – FNMR Versus FMR

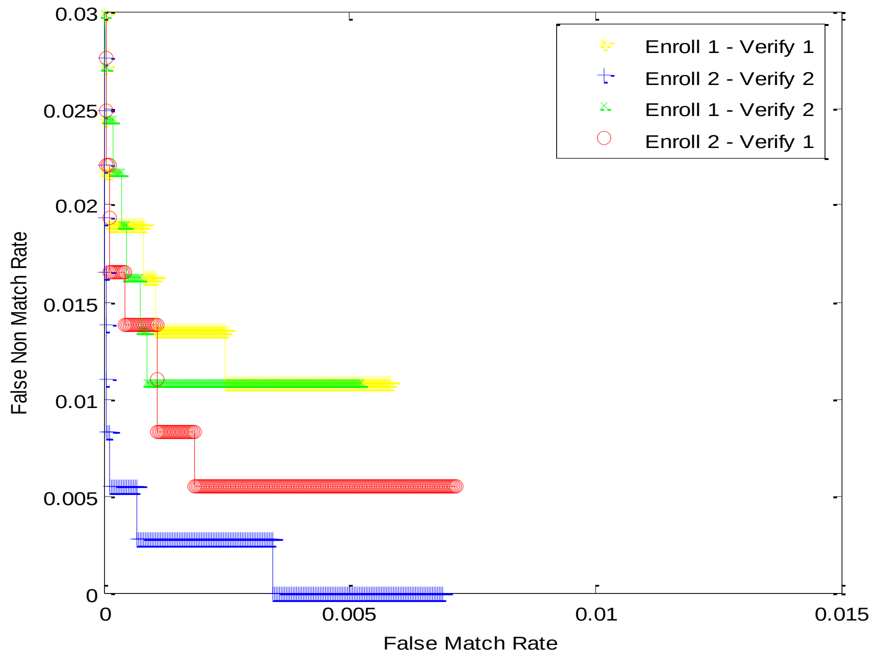


Figure 99. Enroll F Verify A1 – GFRR Versus GFAR

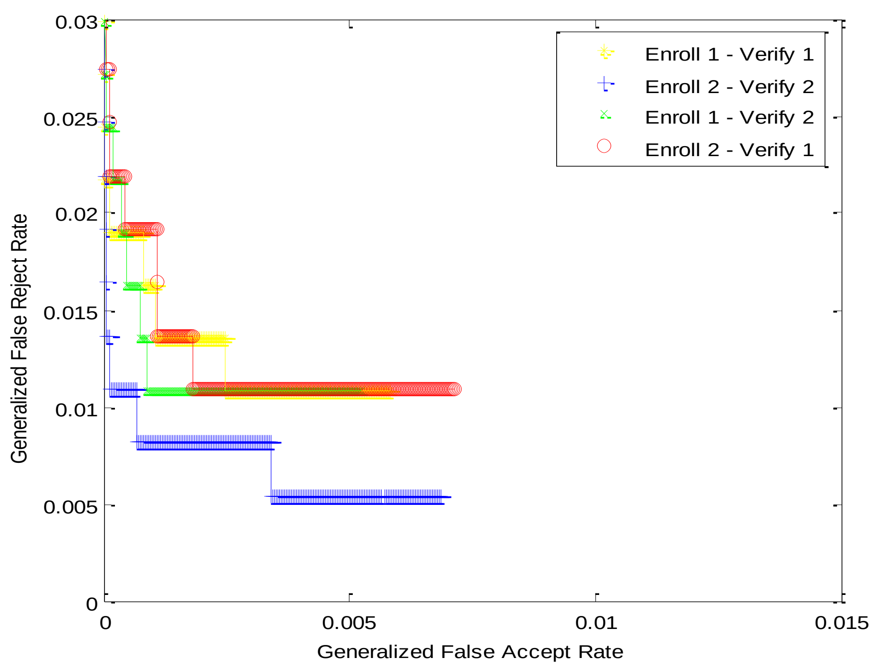


Figure 100. Enroll F Verify B – FNMR Versus FMR

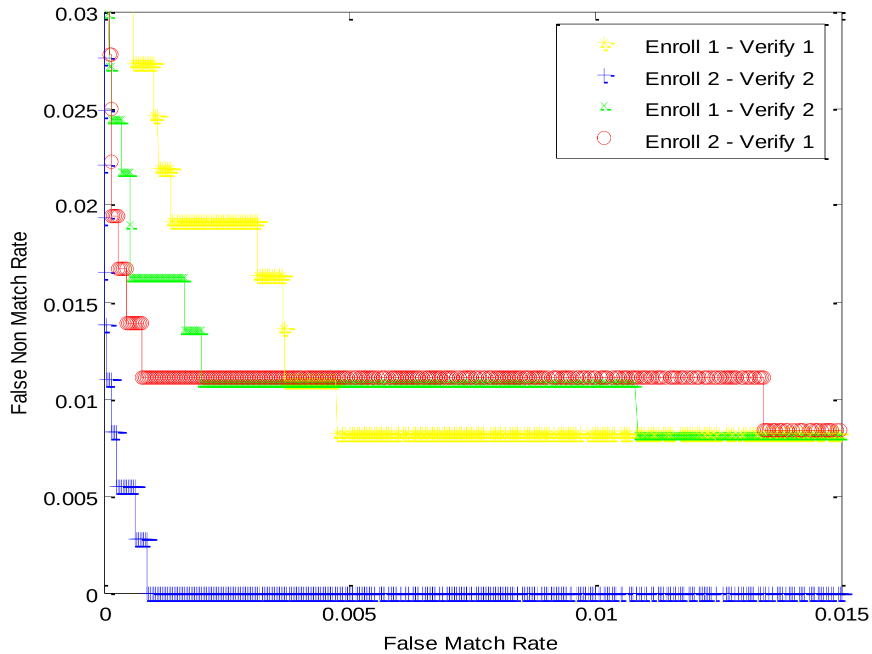


Figure 101. Enroll F Verify B – GFRR Versus GFAR

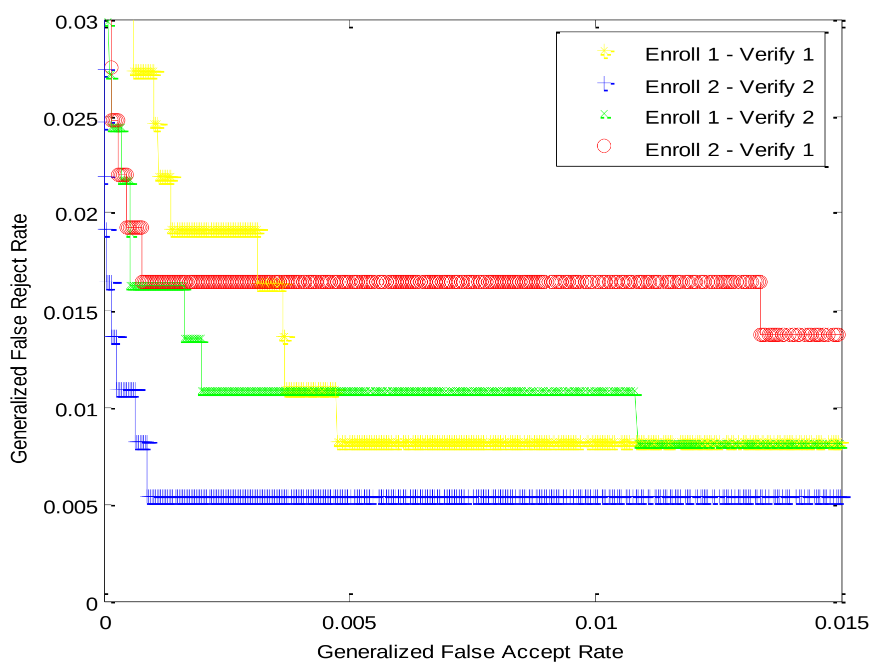


Figure 102. Enroll F Verify C – FNMR Versus FMR

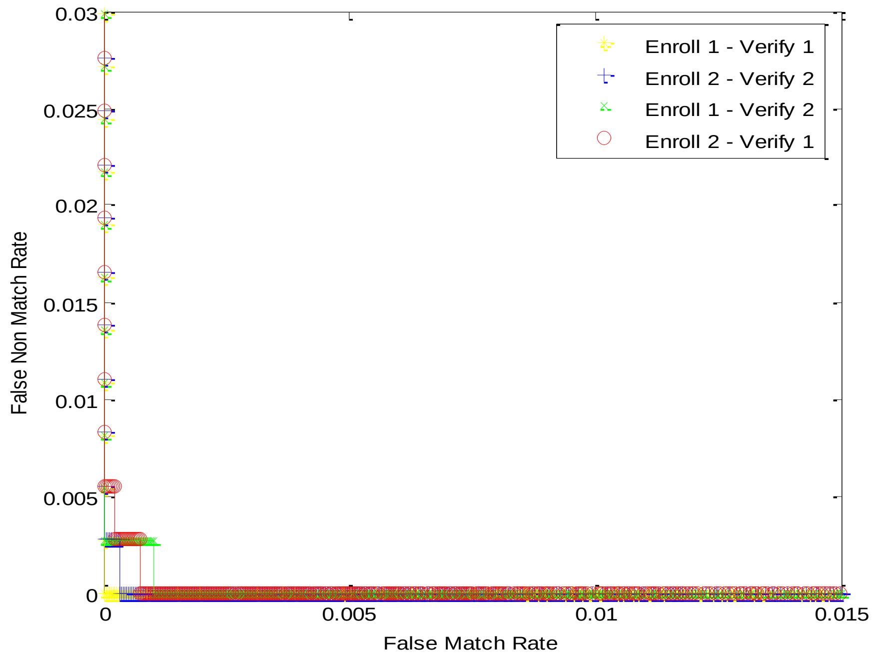


Figure 103. Enroll F Verify C – GFRR Versus GFAR

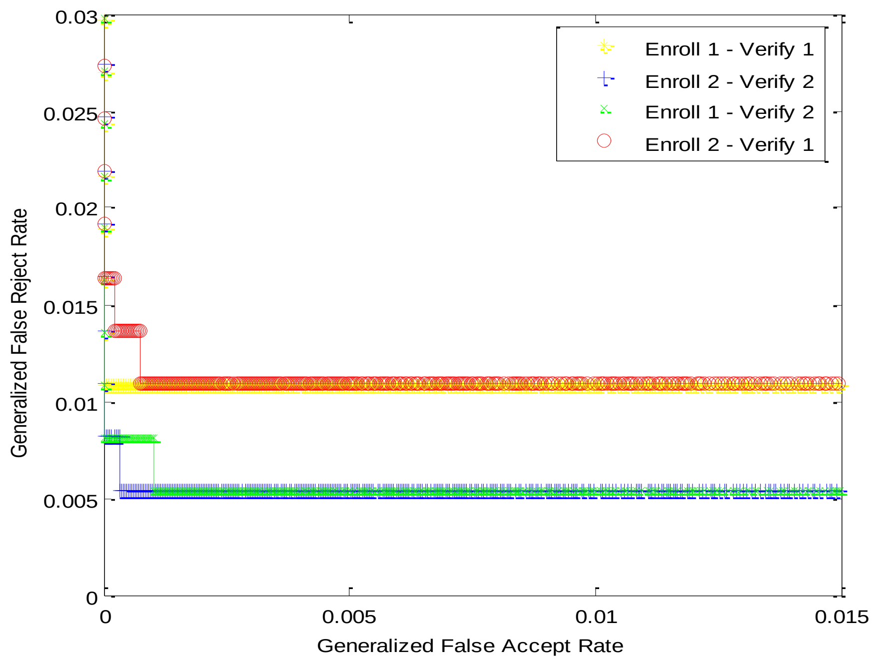


Figure 104. Enroll F Verify D – FNMR Versus FMR

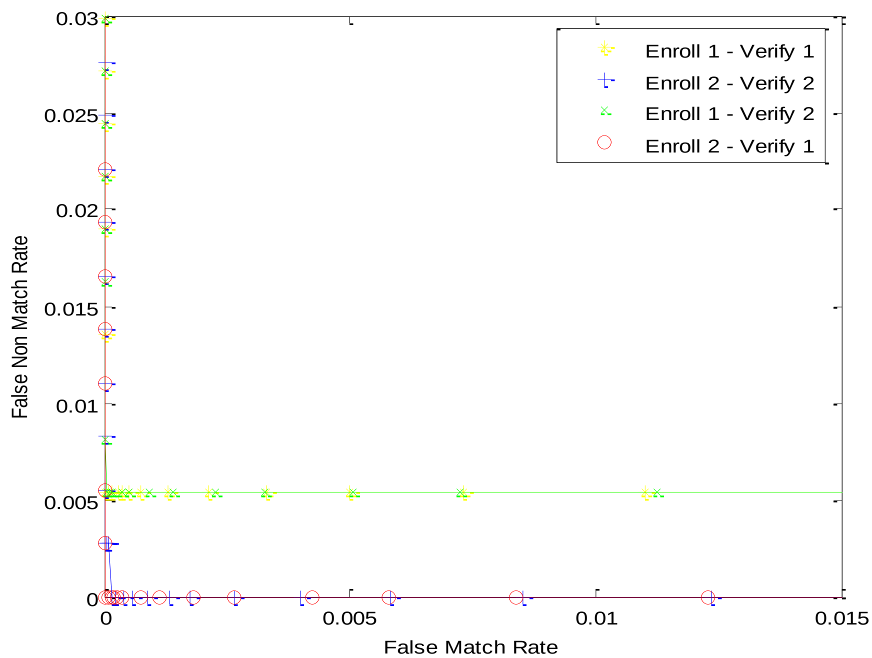


Figure 105. Enroll F Verify D – GFRR Versus GFAR

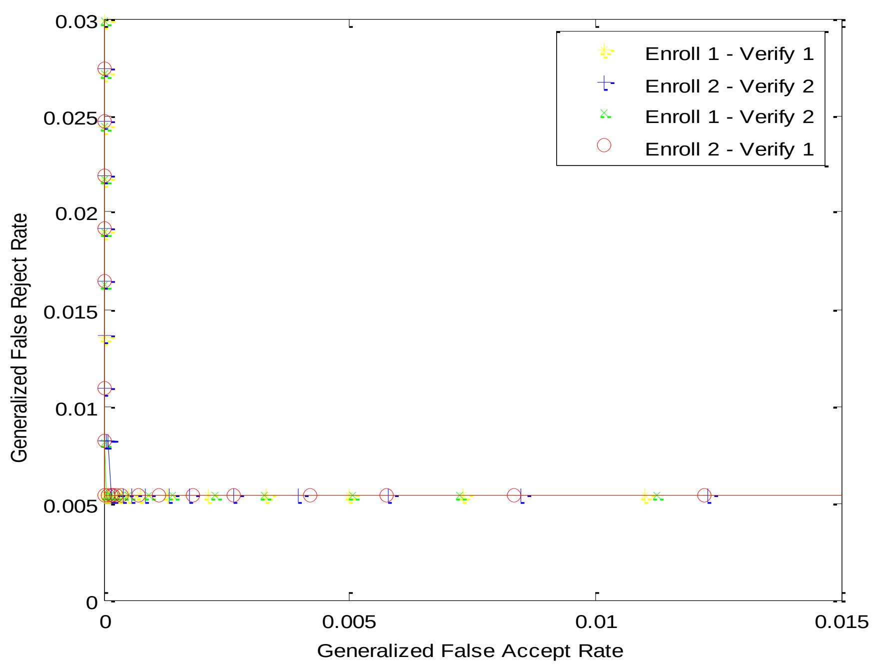


Figure 106. Enroll F Verify E – FNMR Versus FMR

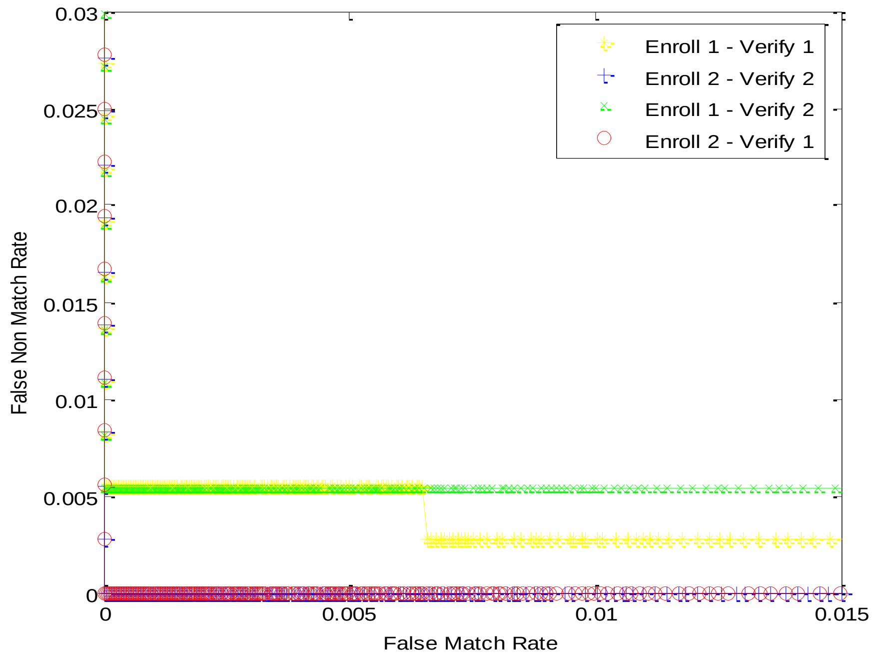


Figure 107. Enroll F Verify E – GFRR Versus GFAR

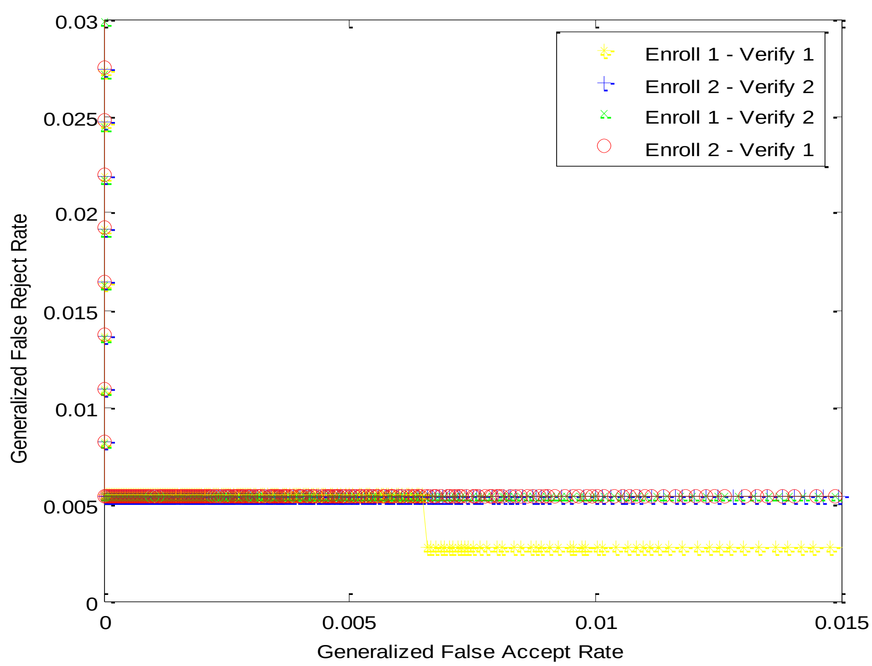


Figure108. Enroll F Verify F – FNMR Versus FMR

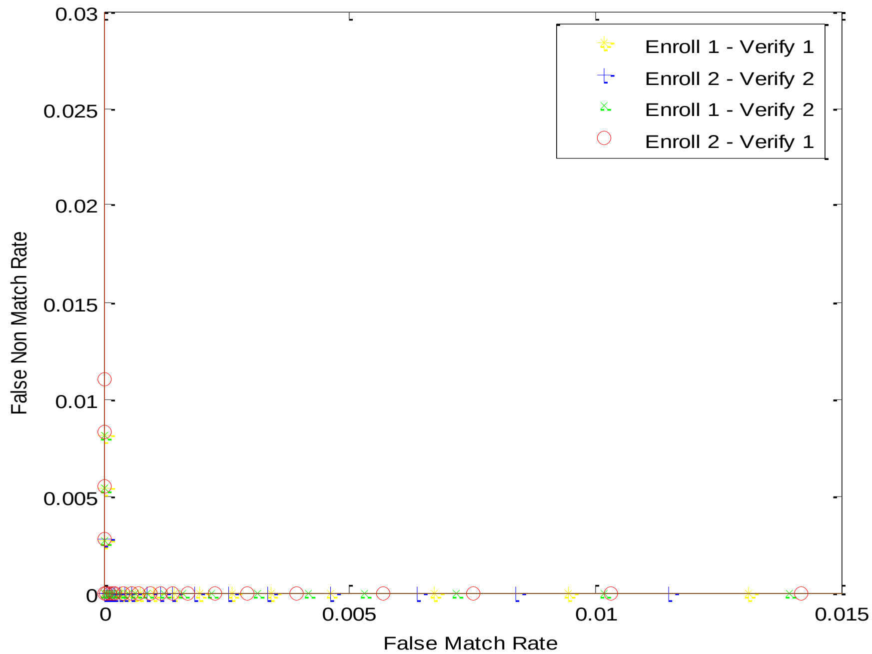


Figure 109. Enroll F Verify F – GFRR Versus GFAR

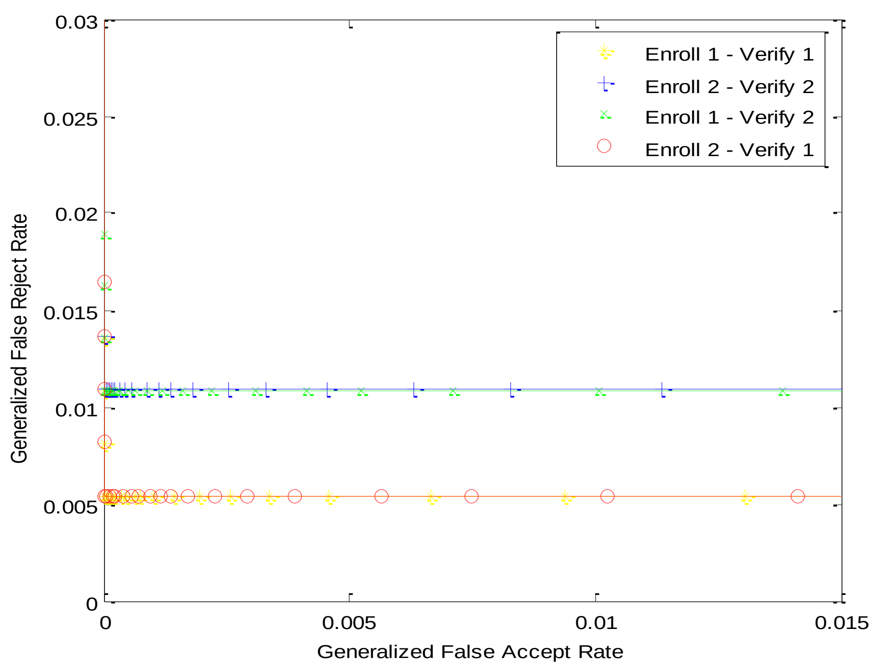


Figure 110. Enroll F Verify G – FNMR Versus FMR

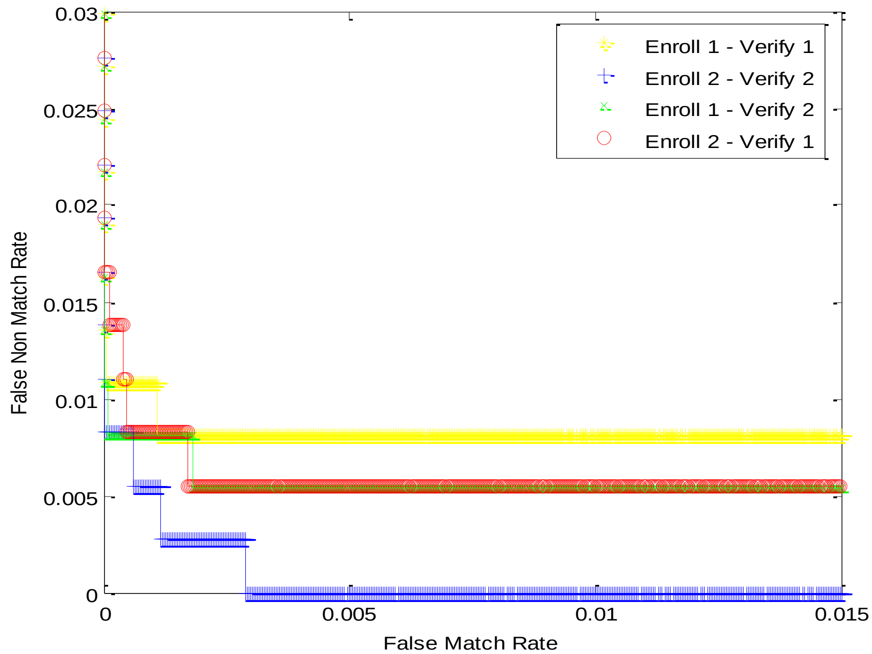


Figure 111. Enroll F Verify G – GFRR Versus GFAR

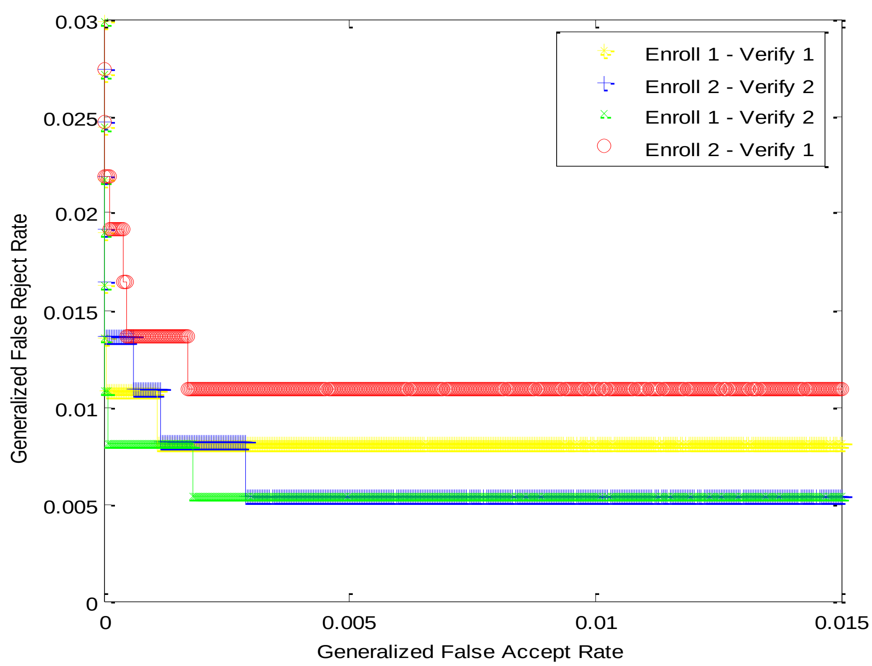


Figure 112. Enroll F Verify H – FNMR Versus FMR

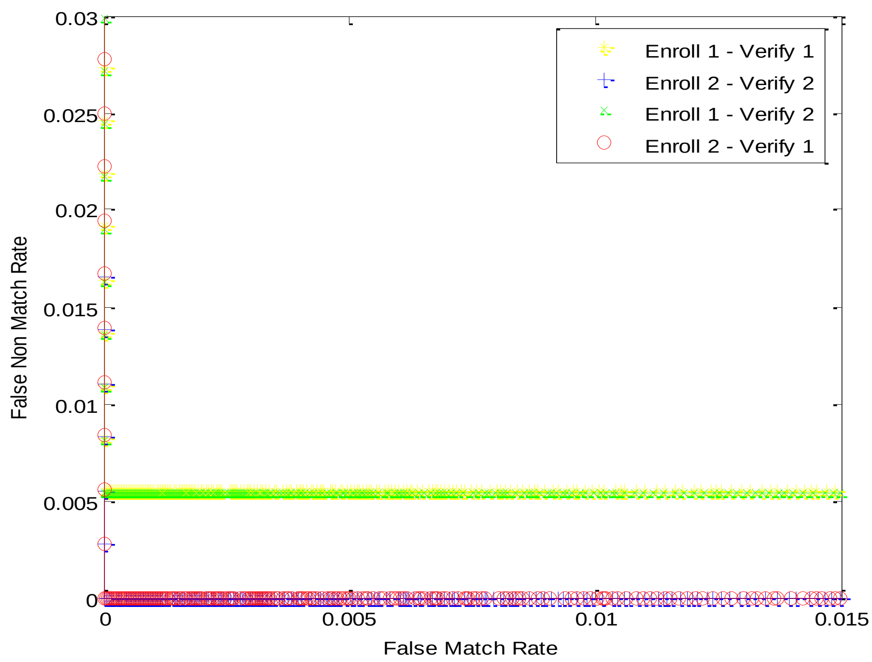


Figure 113. Enroll F Verify H – GFRR Versus GFAR

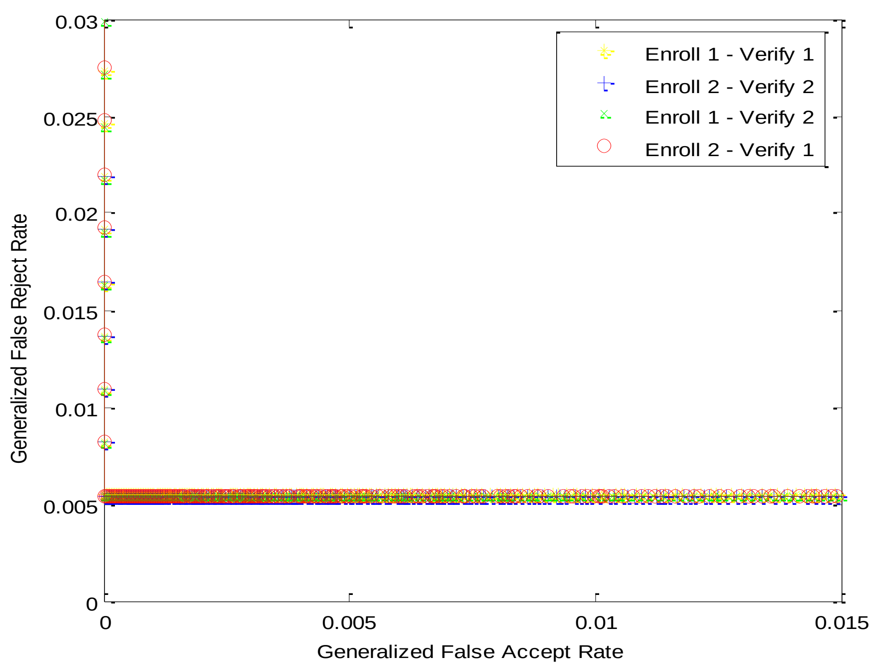


Figure 114. Enroll F Verify I – FNMR Versus FMR

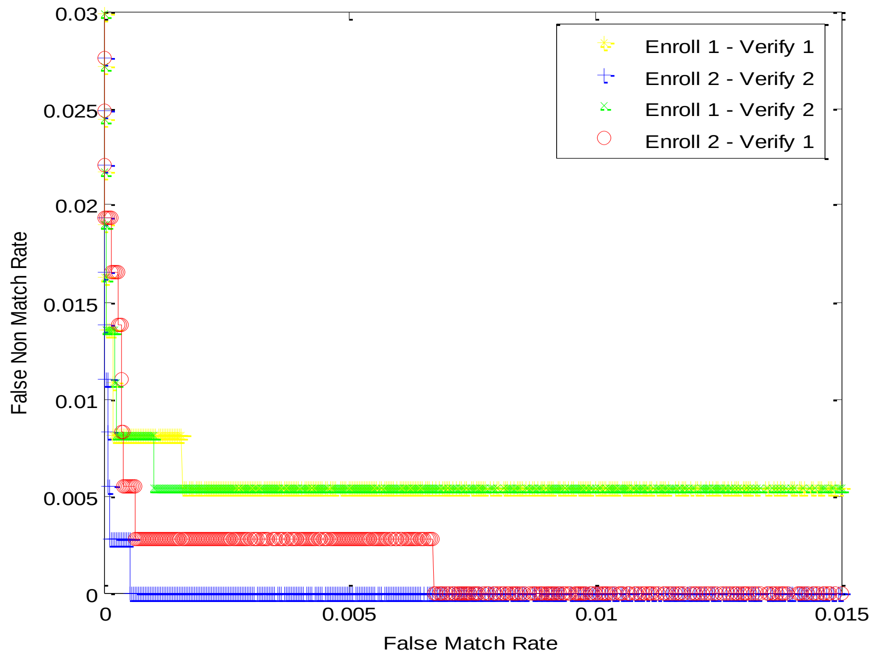


Figure 115. Enroll F Verify I – GFRR Versus GFAR

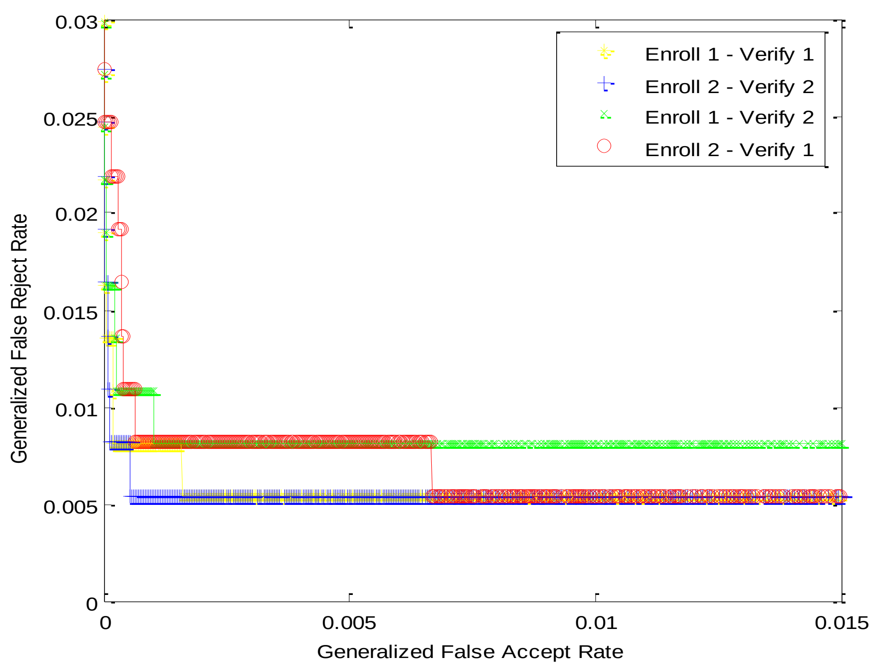


Figure 116. Enroll G Verify A1 – FNMR Versus FMR

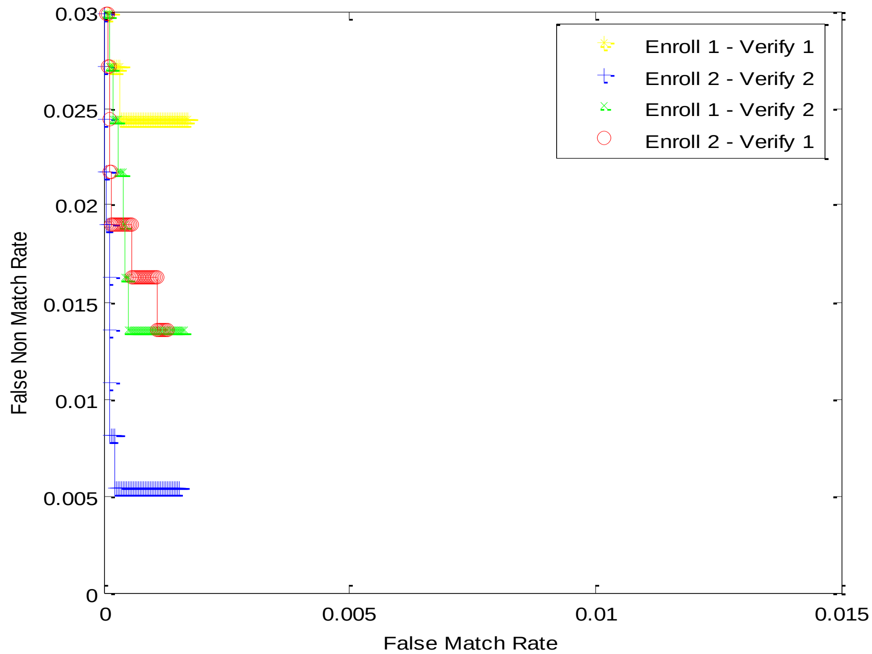


Figure 117. Enroll G Verify A1 – GFRR Versus GFAR

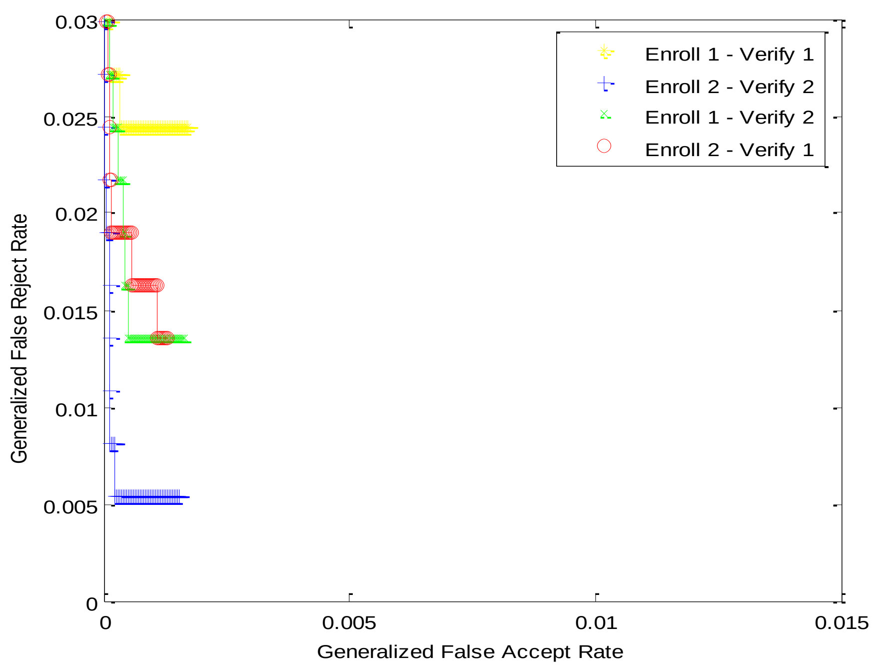


Figure 118. Enroll G Verify B – FNMR Versus FMR

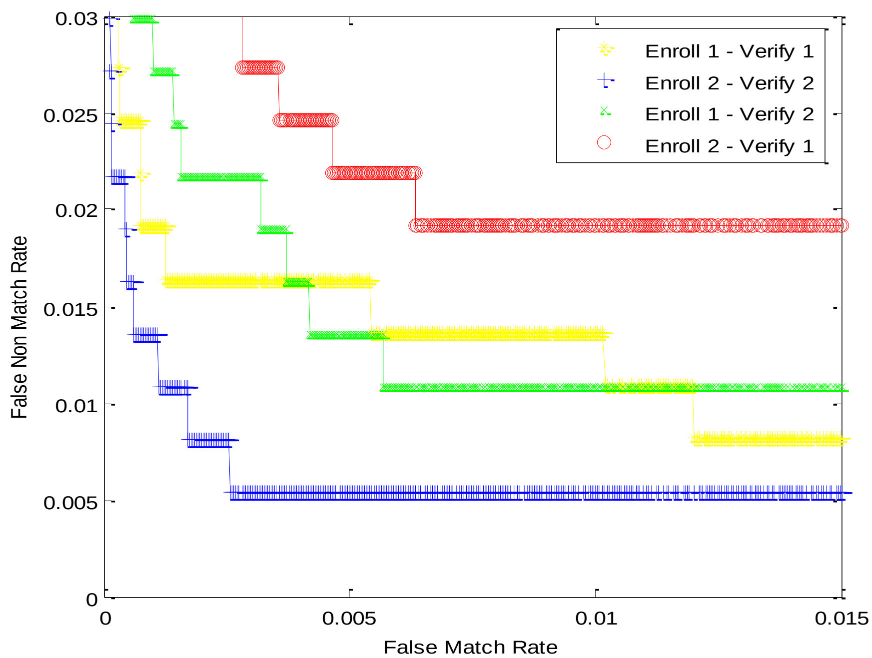


Figure 119. Enroll G Verify B – GFRR Versus GFAR

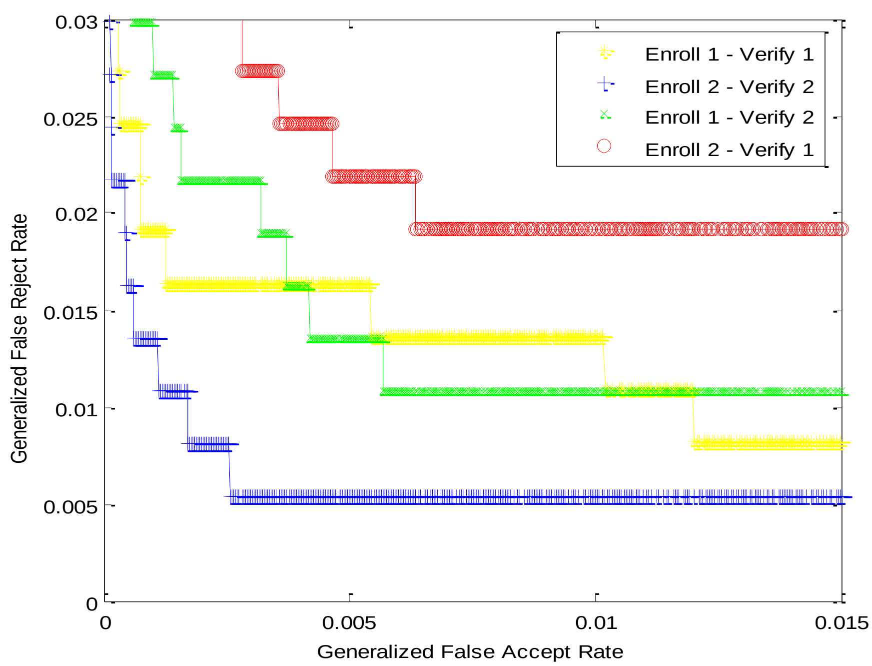


Figure 120. Enroll G Verify C – FNMR Versus FMR

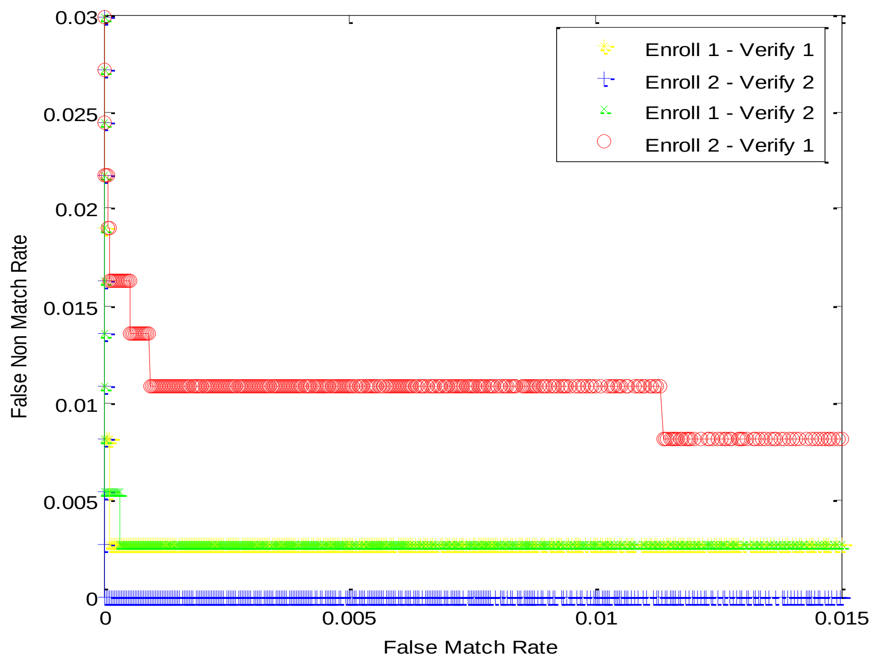


Figure 121. Enroll G Verify C – GFRR Versus GFAR

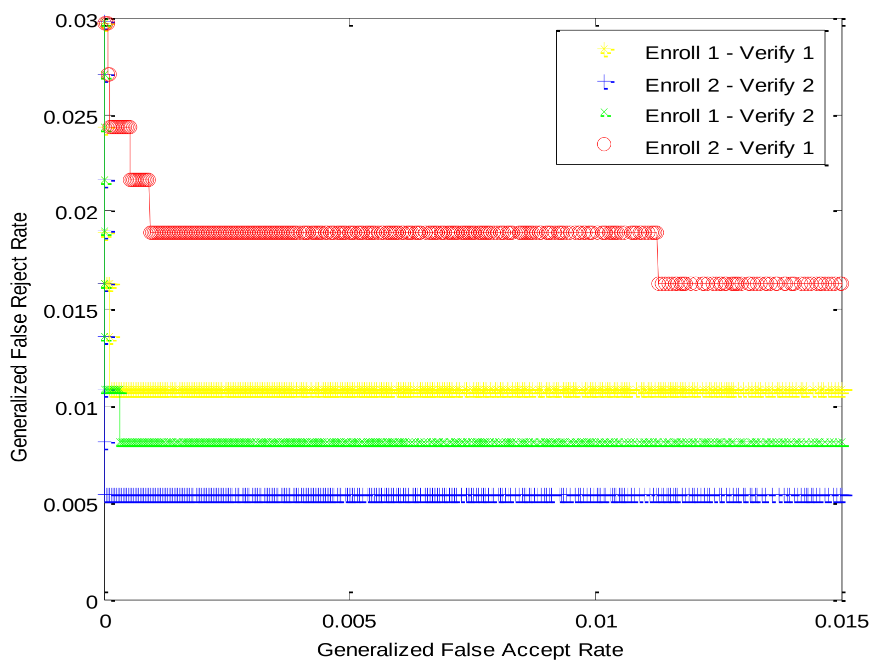


Figure 122. Enroll G Verify D – FNMR Versus FMR

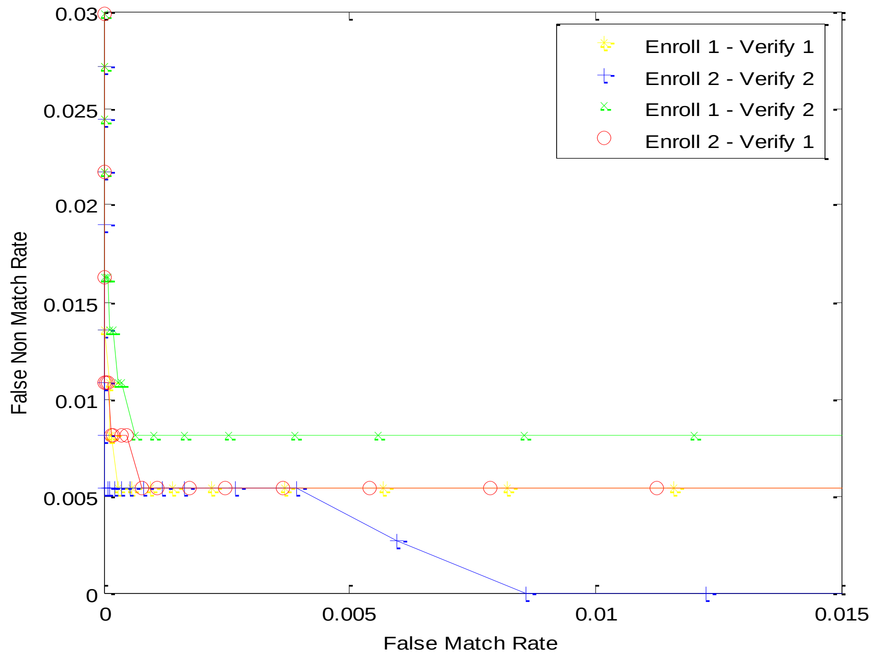


Figure 123. Enroll G Verify D – GFRR Versus GFAR

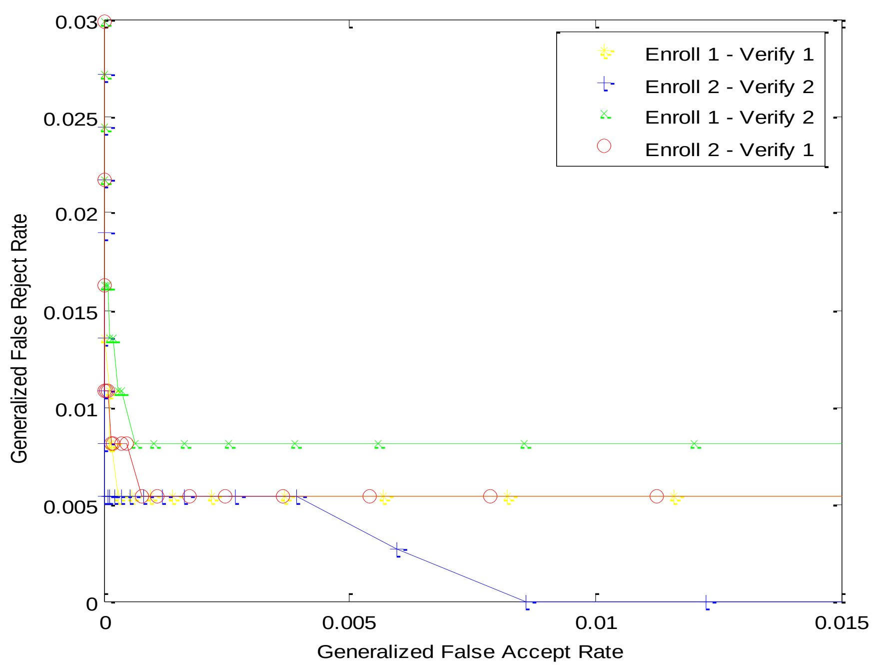


Figure 124. Enroll G Verify E – FNMR Versus FMR

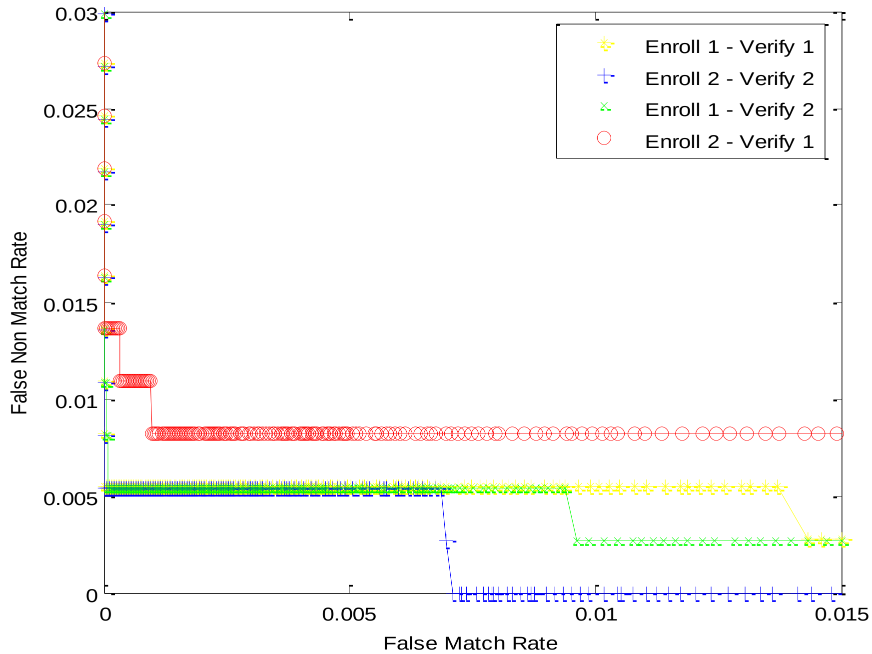


Figure 125. Enroll G Verify E – GFRR Versus GFAR

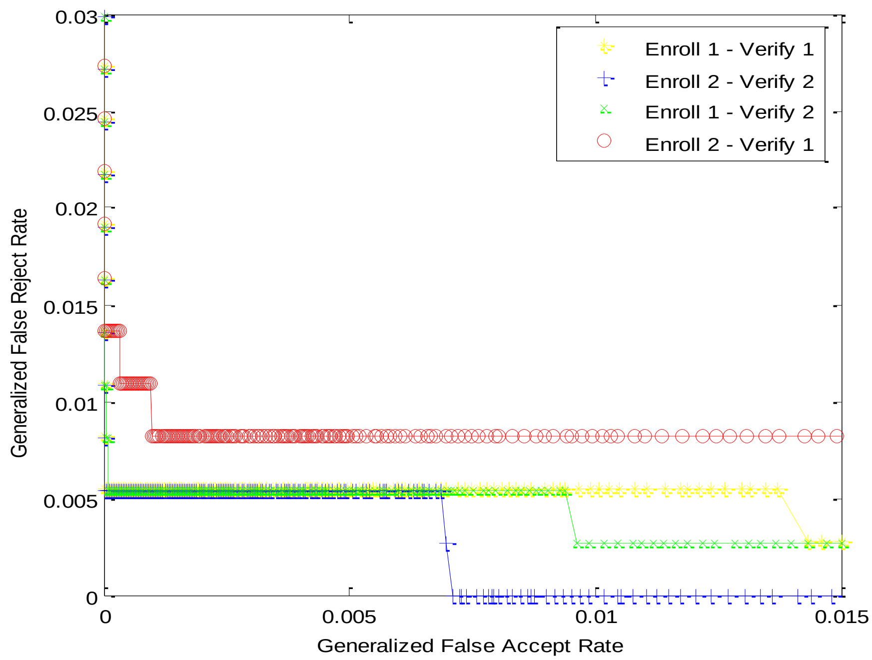


Figure 126. Enroll G Verify F – FNMR Versus FMR

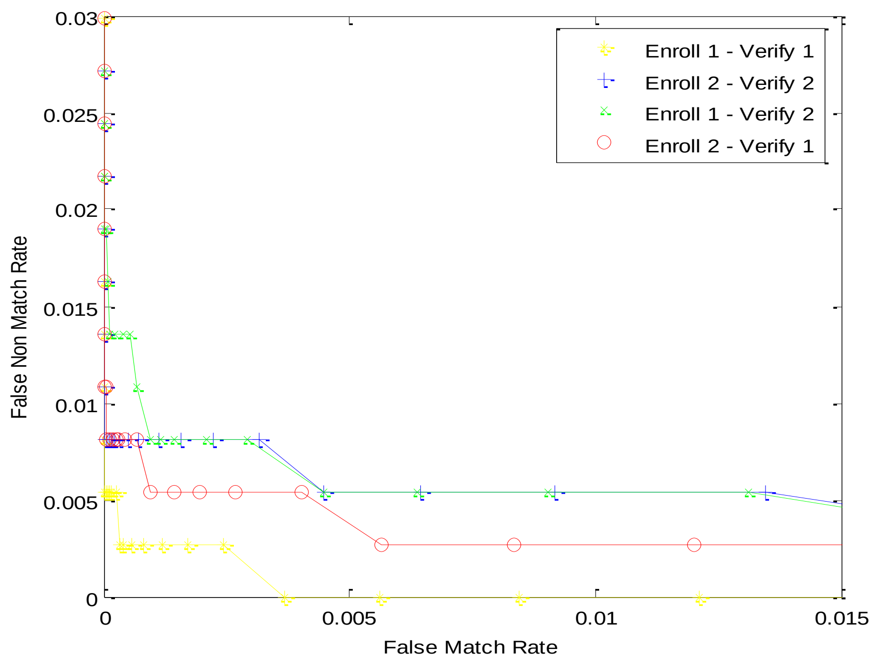


Figure 127. Enroll G Verify F – GFRR Versus GFAR

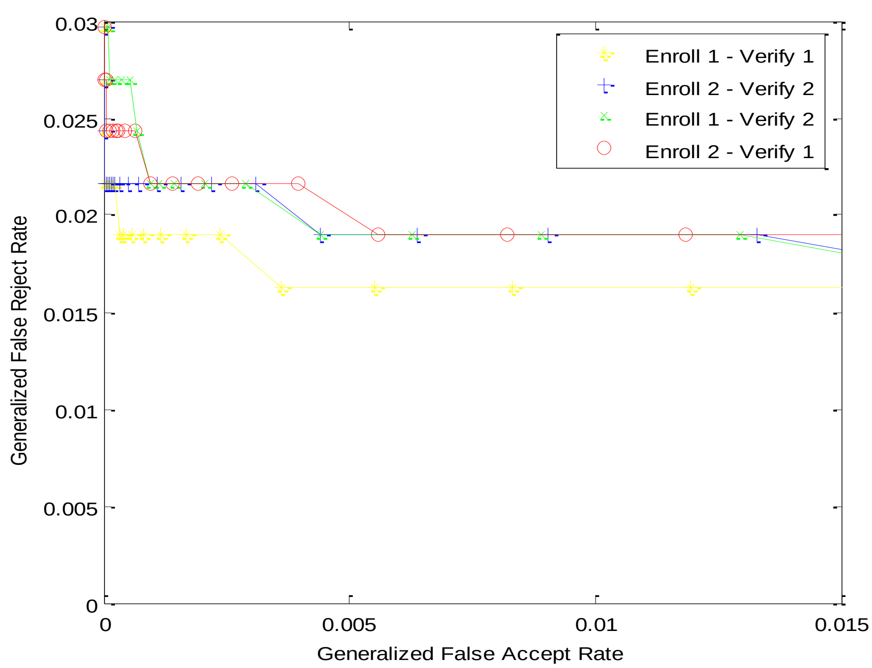


Figure 128. Enroll G Verify G – FNMR Versus FMR

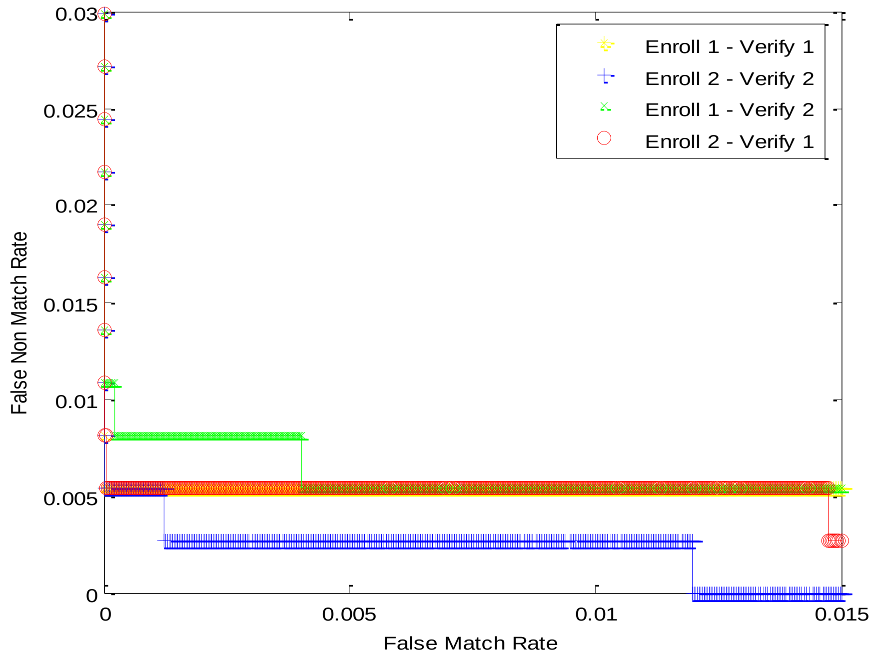


Figure 129. Enroll G Verify G – GFRR Versus GFAR

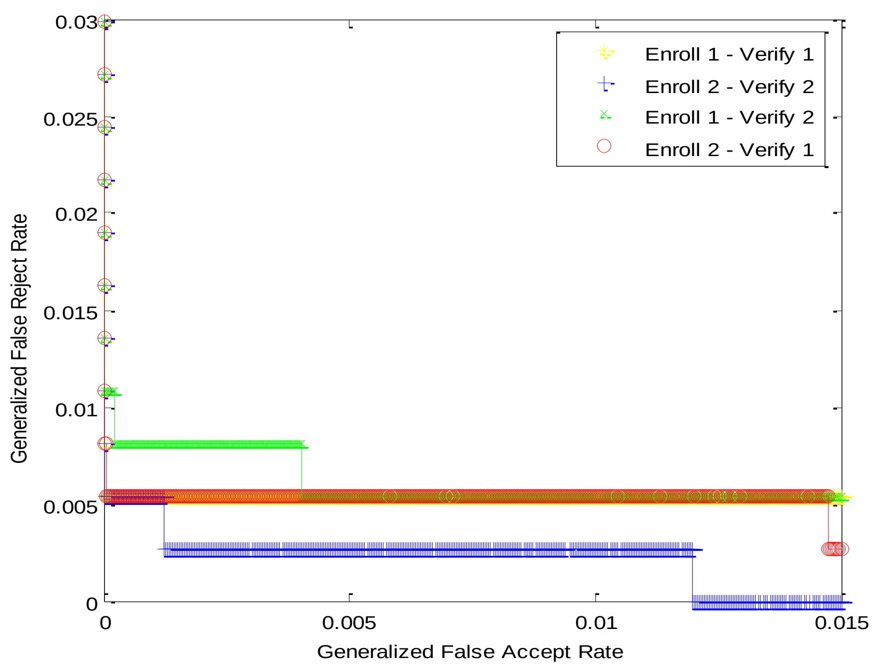


Figure 130. Enroll G Verify H – FNMR Versus FMR

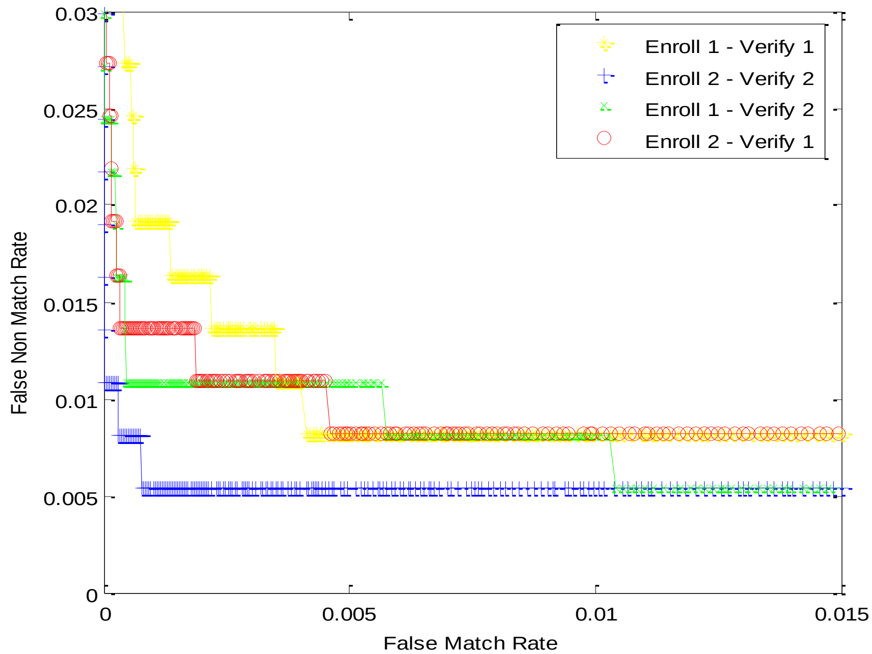


Figure 131. Enroll G Verify H – GFRR Versus GFAR

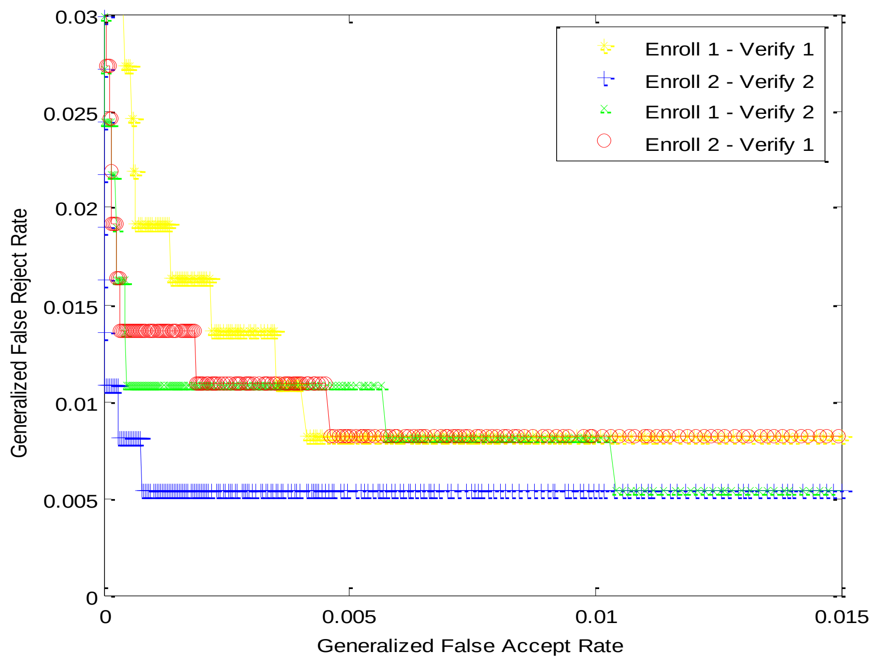


Figure 132. Enroll G Verify I – FNMR Versus FMR

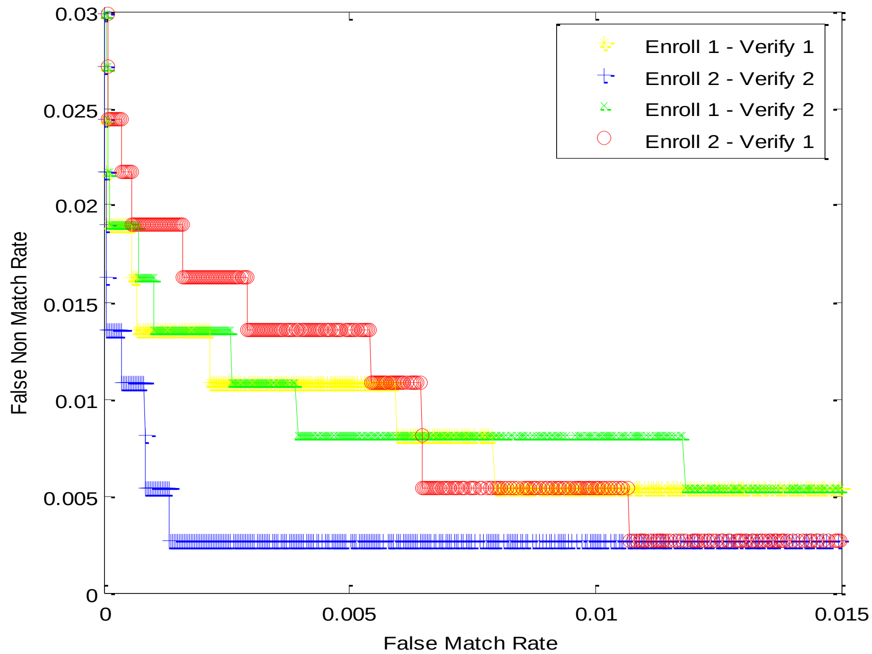


Figure 133. Enroll G Verify I – GFRR Versus GFAR

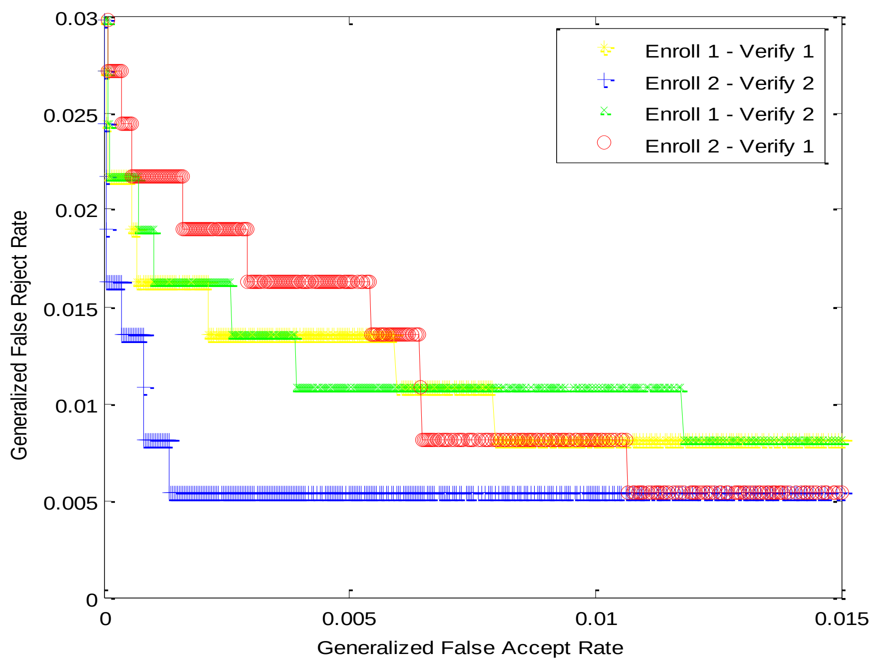


Figure 134. Enroll H Verify A1 – FNMR Versus FMR

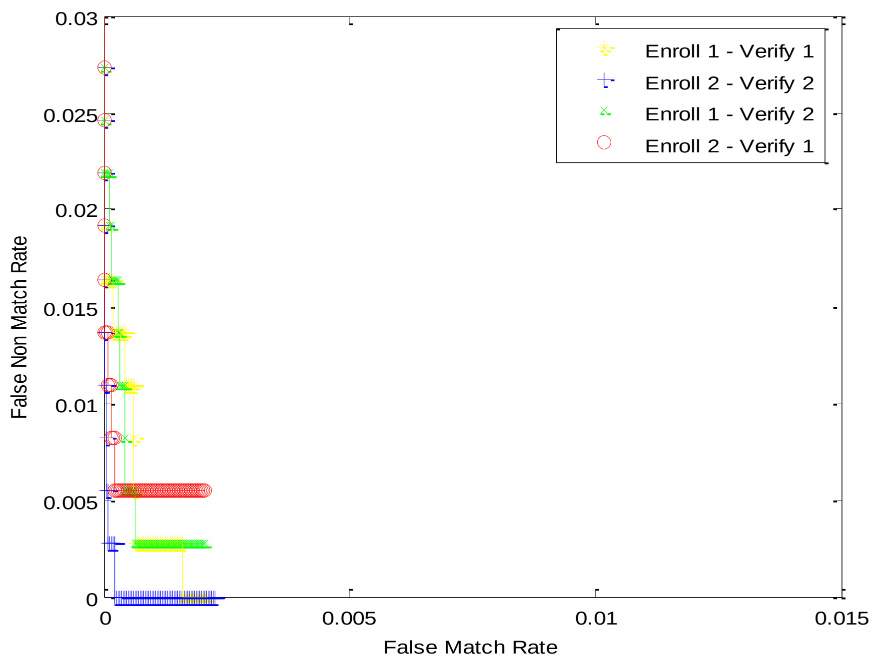


Figure 135. Enroll H Verify A1 – GFRR Versus GFAR

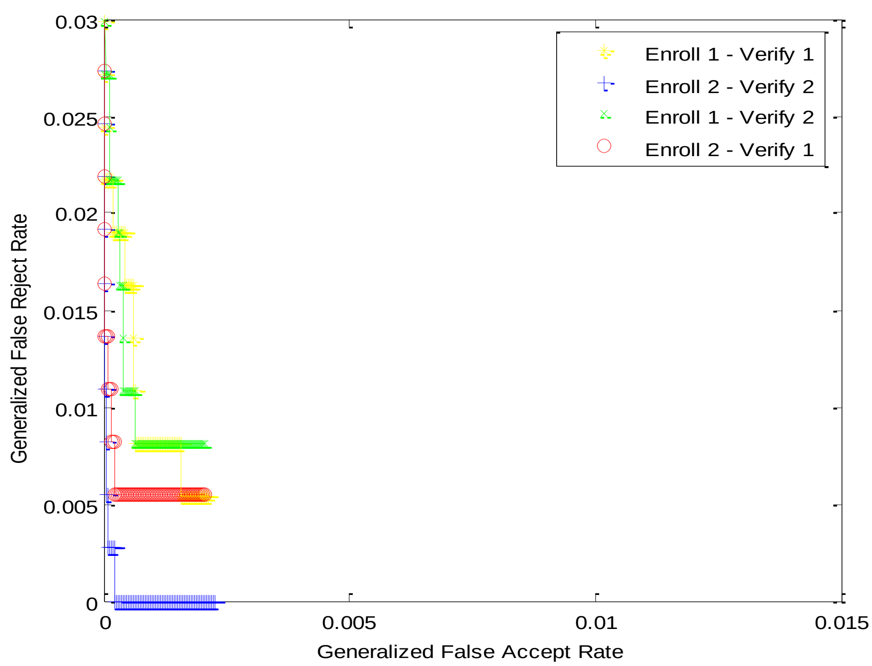


Figure 136. Enroll H Verify B – FNMR Versus FMR

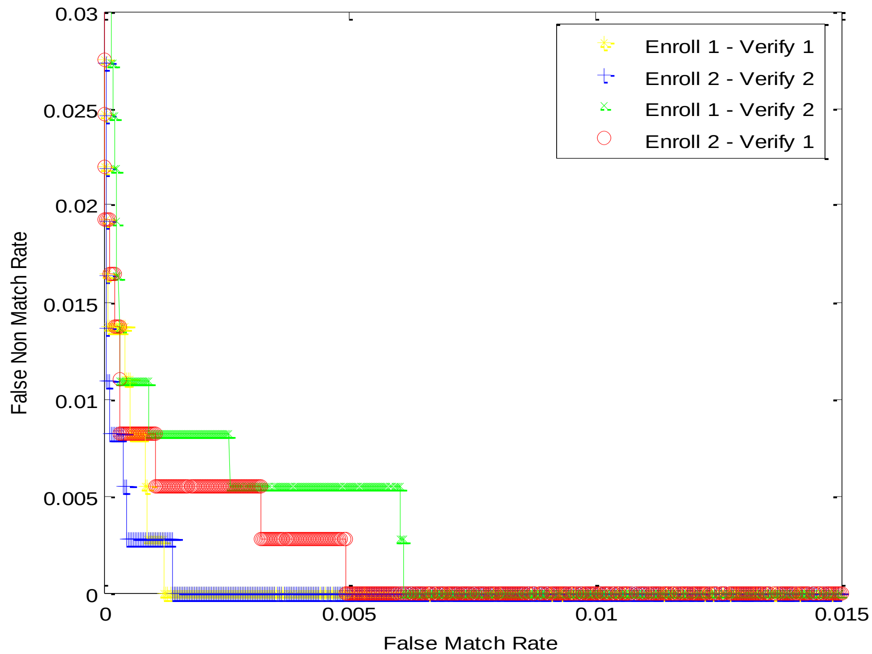


Figure 137. Enroll H Verify B – GFRR Versus GFAR

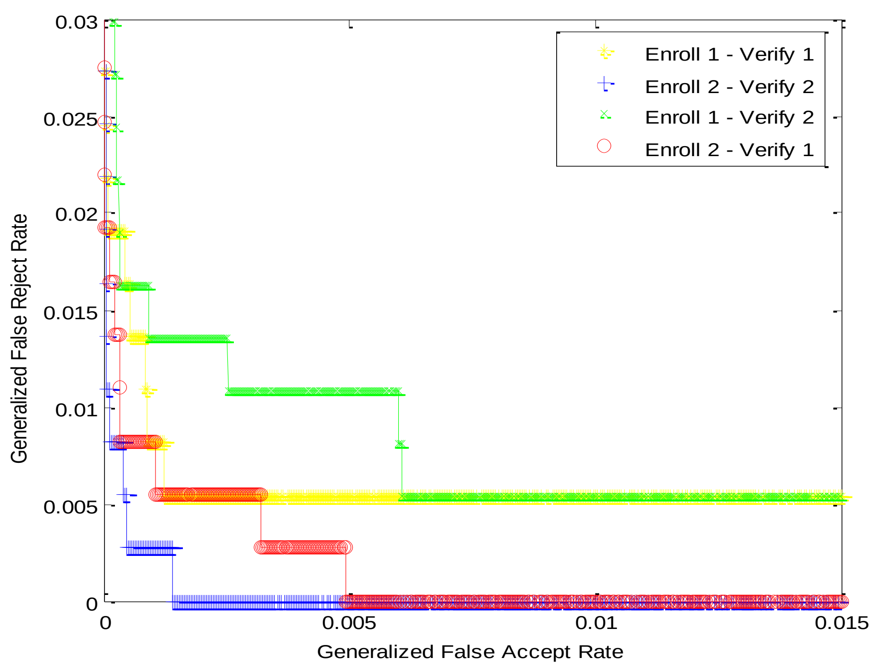


Figure 138. Enroll H Verify C – FNMR Versus FMR

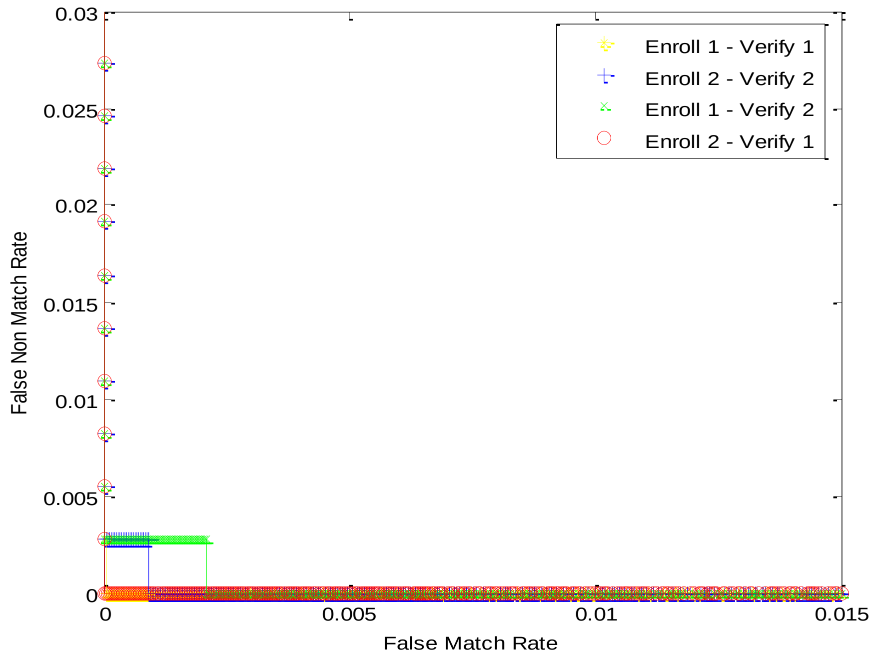


Figure 139. Enroll H Verify C – GFRR Versus GFAR

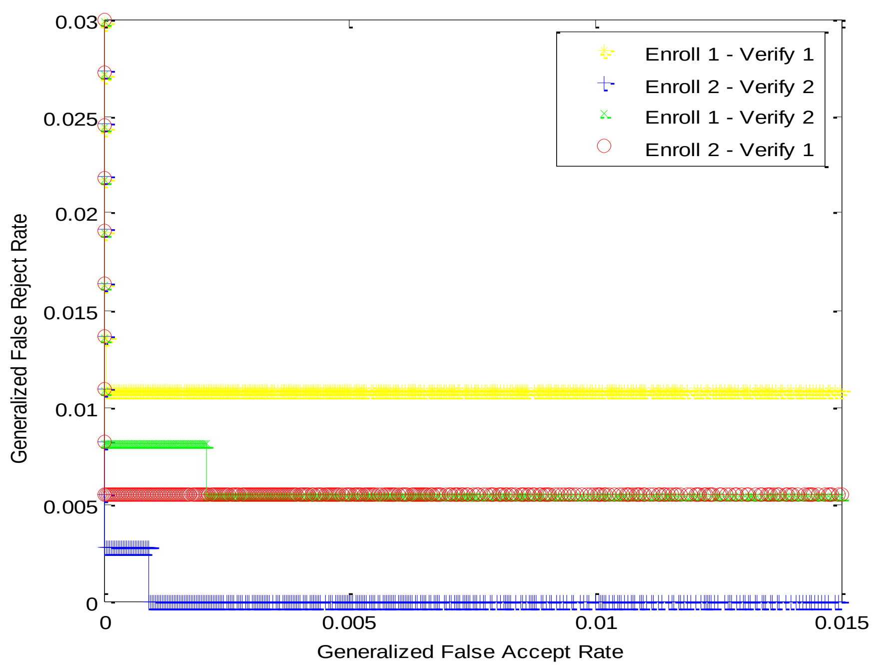


Figure 140. Enroll H Verify D – FNMR Versus FMR

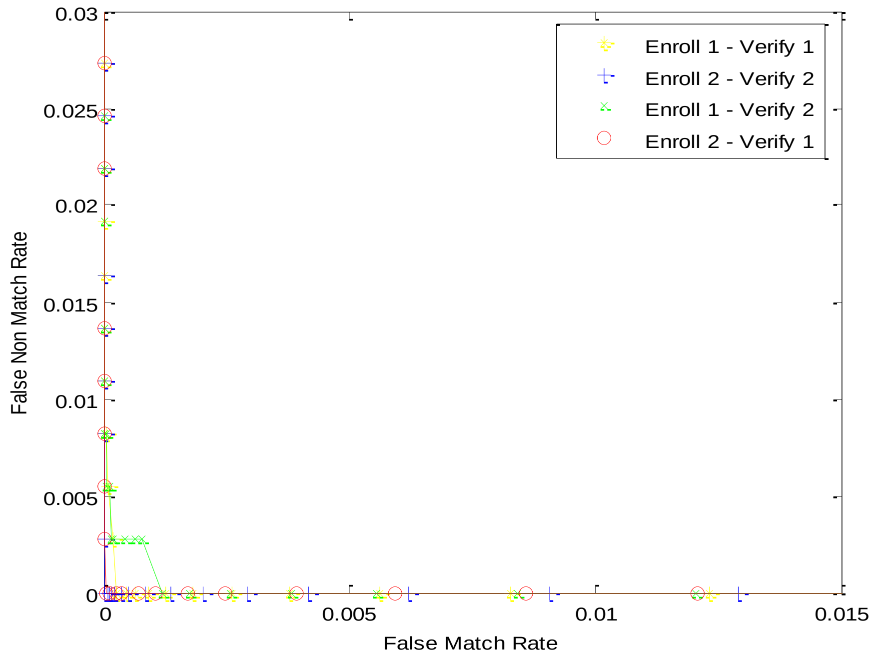


Figure 141. Enroll H Verify D – GFRR Versus GFAR

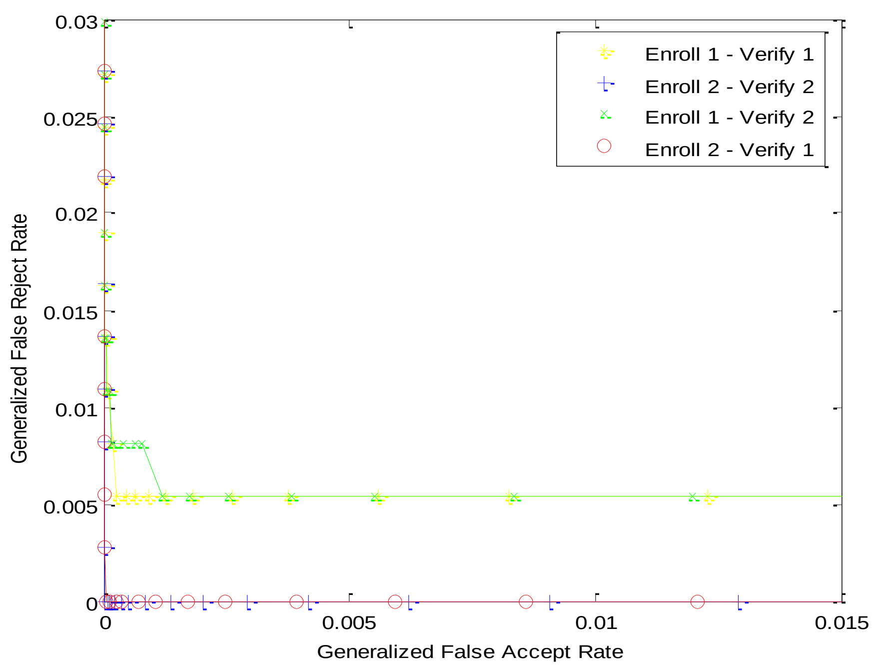


Figure 142. Enroll H Verify E – FNMR Versus FMR

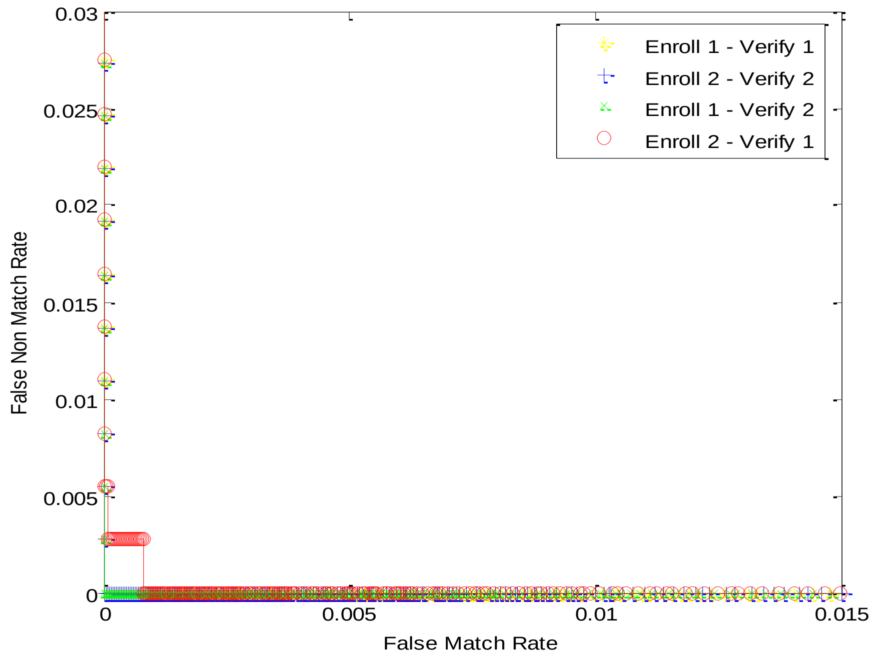


Figure 143. Enroll H Verify E – GFRR Versus GFAR

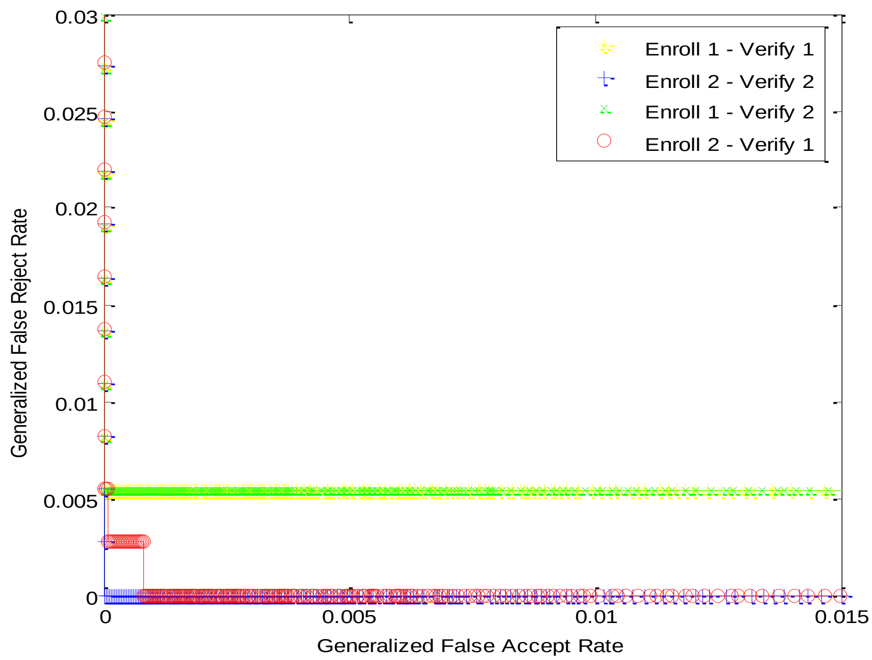


Figure 144. Enroll H Verify F – FNMR Versus FMR

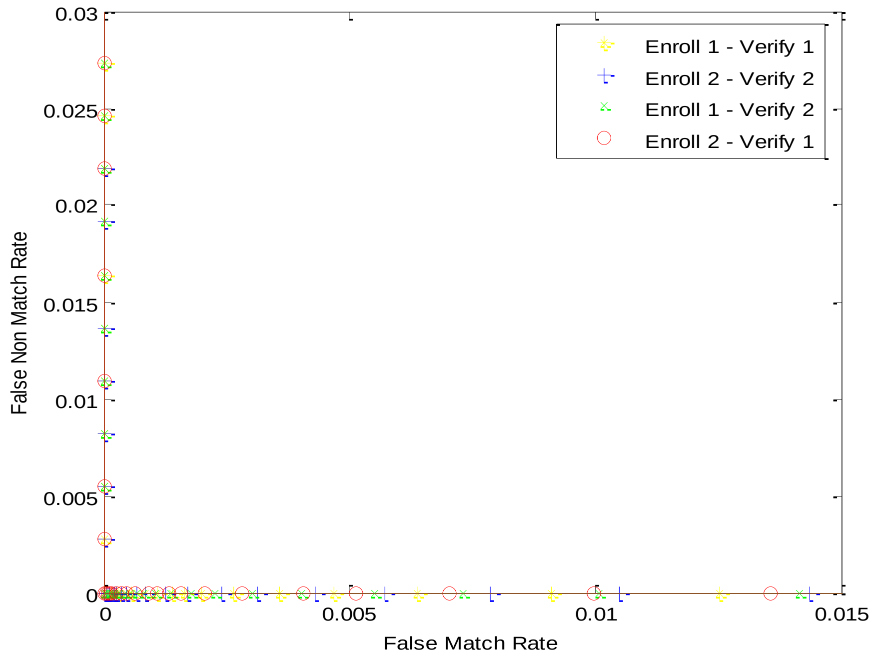


Figure 145. Enroll H Verify F – GFRR Versus GFAR

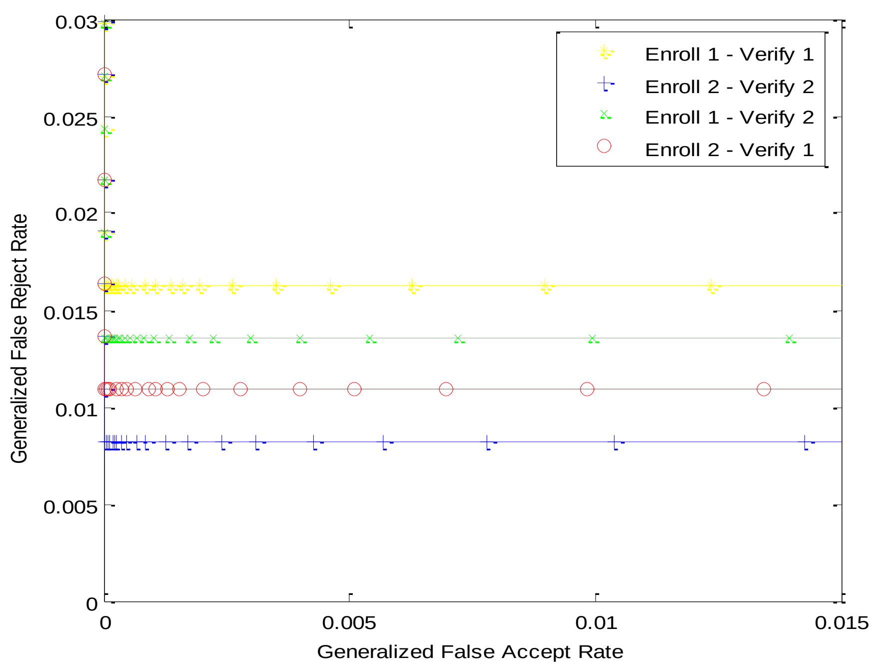


Figure 146. Enroll H Verify G – FNMR Versus FMR

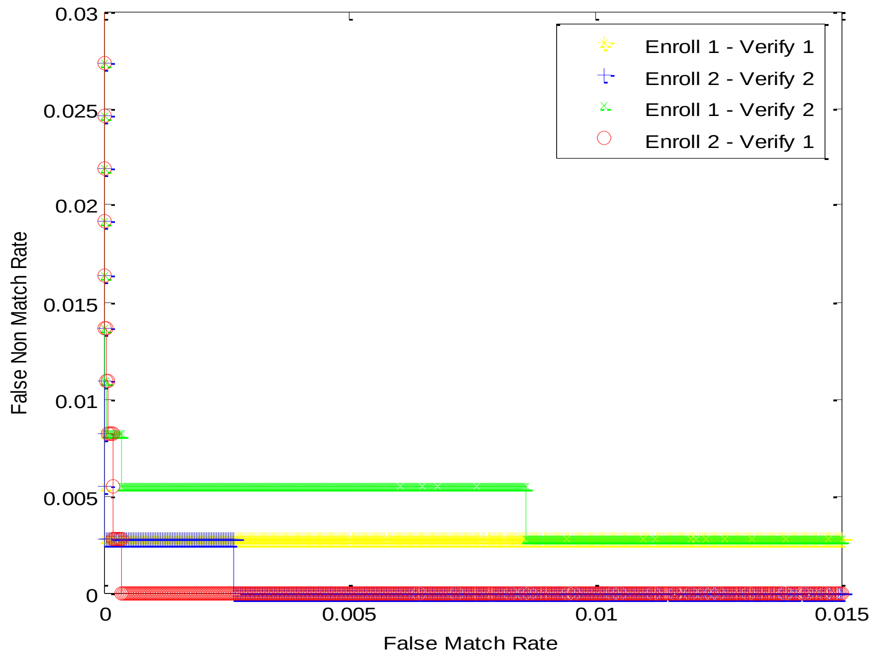


Figure 147. Enroll H Verify G – GFRR Versus GFAR

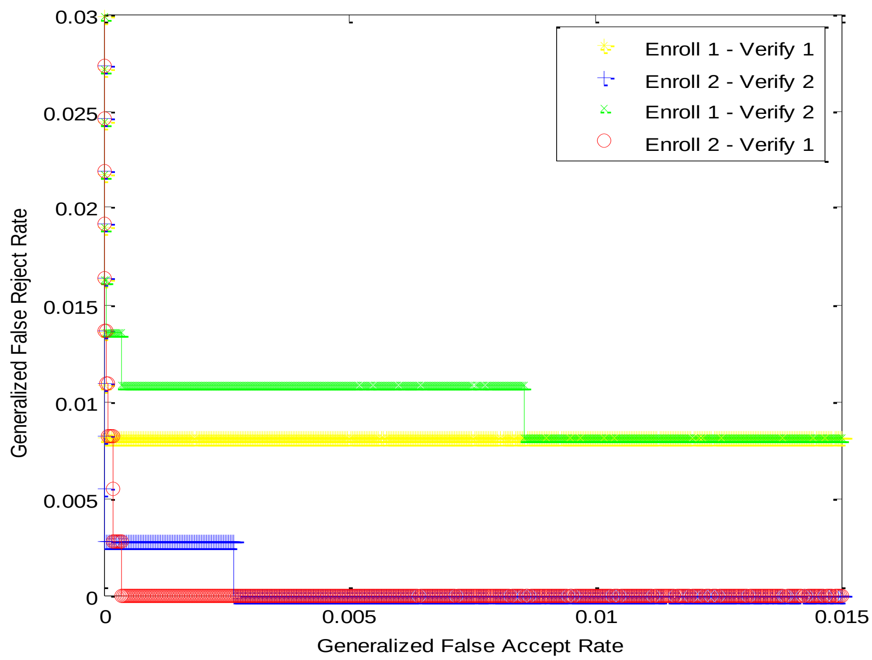


Figure 148. Enroll H Verify H – FNMR Versus FMR

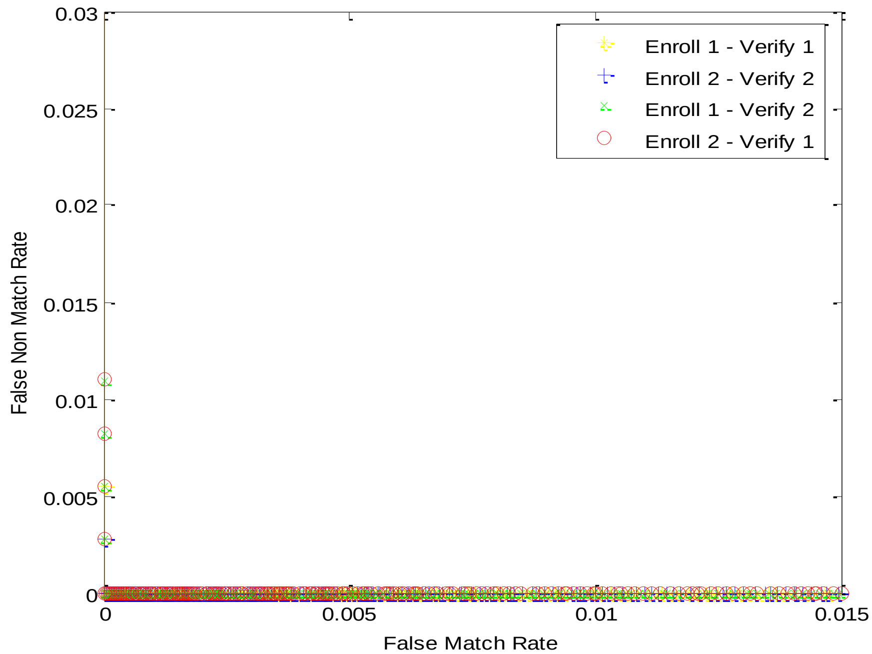


Figure 149. Enroll H Verify H – GFRR Versus GFAR

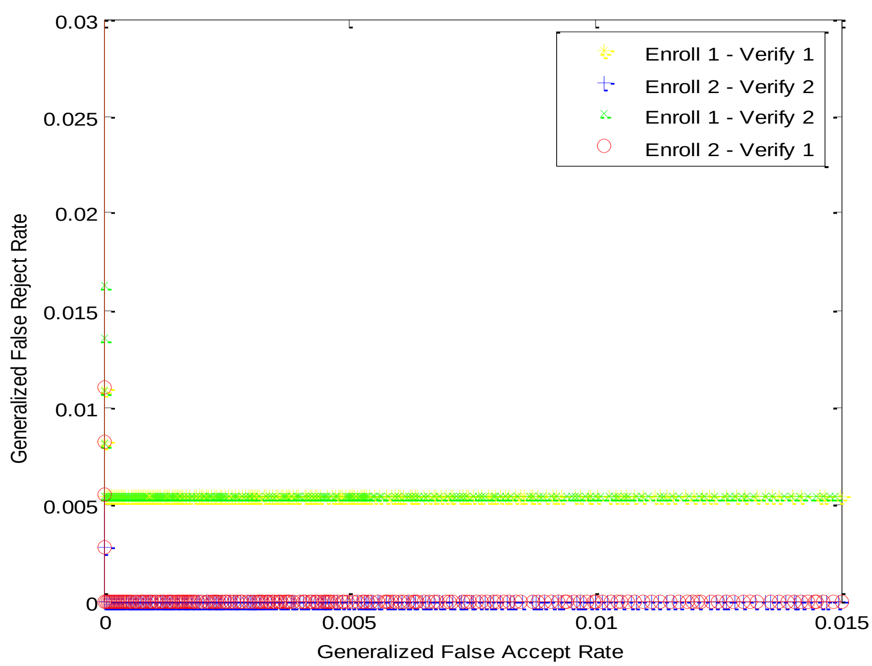


Figure 150. Enroll H Verify I – FNMR Versus FMR

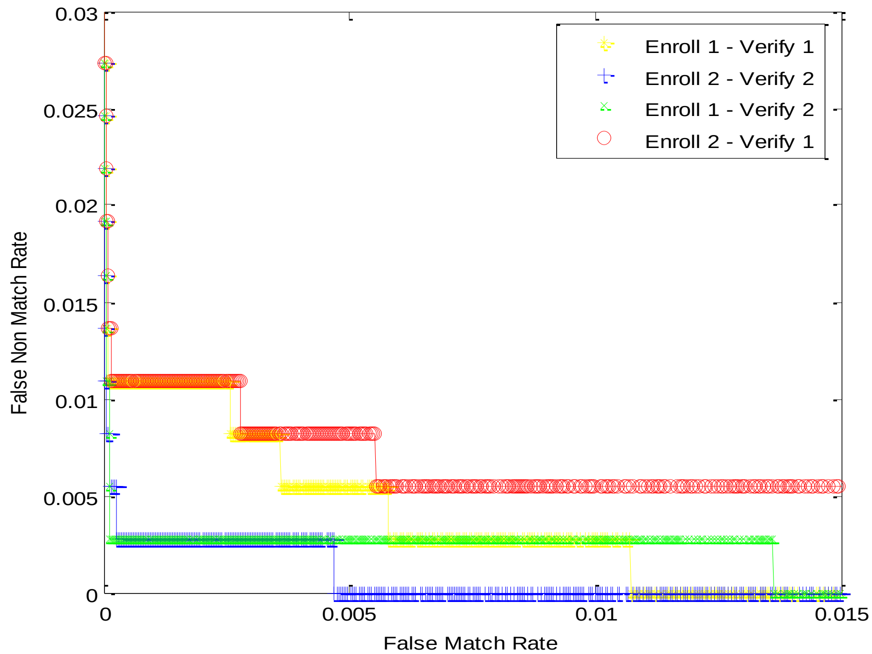


Figure 151. Enroll H Verify I – GFRR Versus GFAR

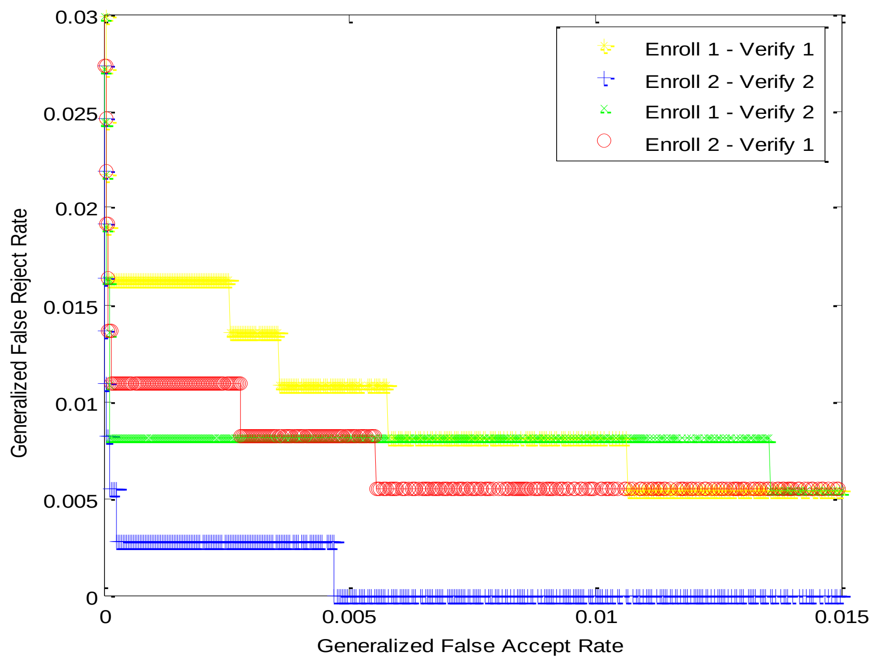


Figure 152. Enroll I Verify A1 – FNMR Versus FMR

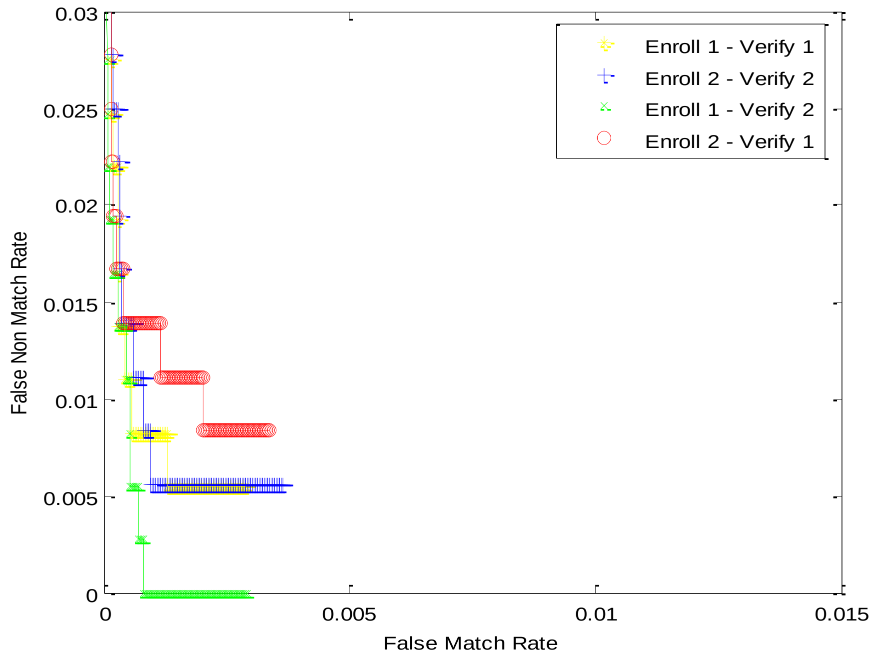


Figure 153. Enroll I Verify A1 – GFRR Versus GFAR

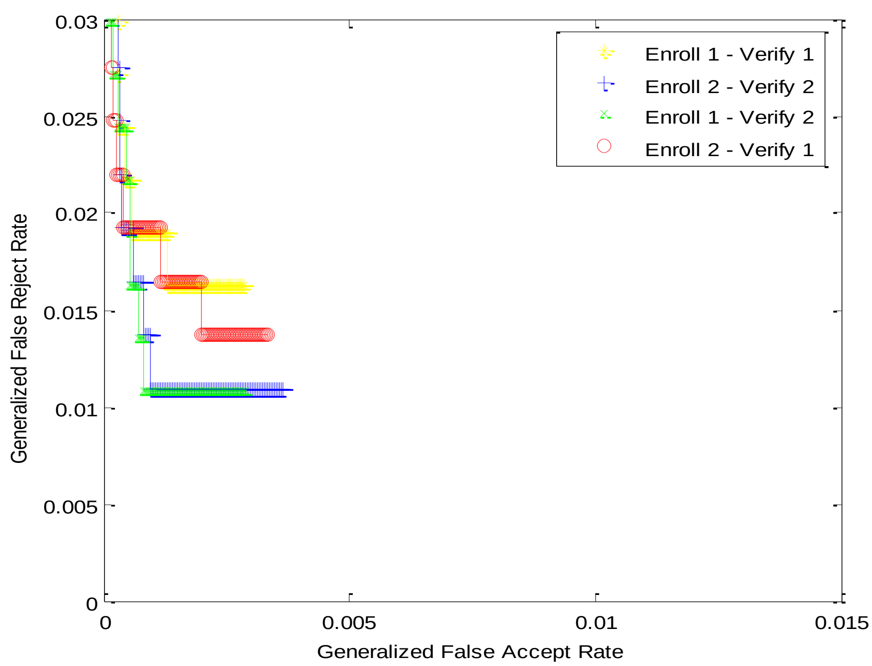


Figure 154. Enroll I Verify B – FNMR Versus FMR

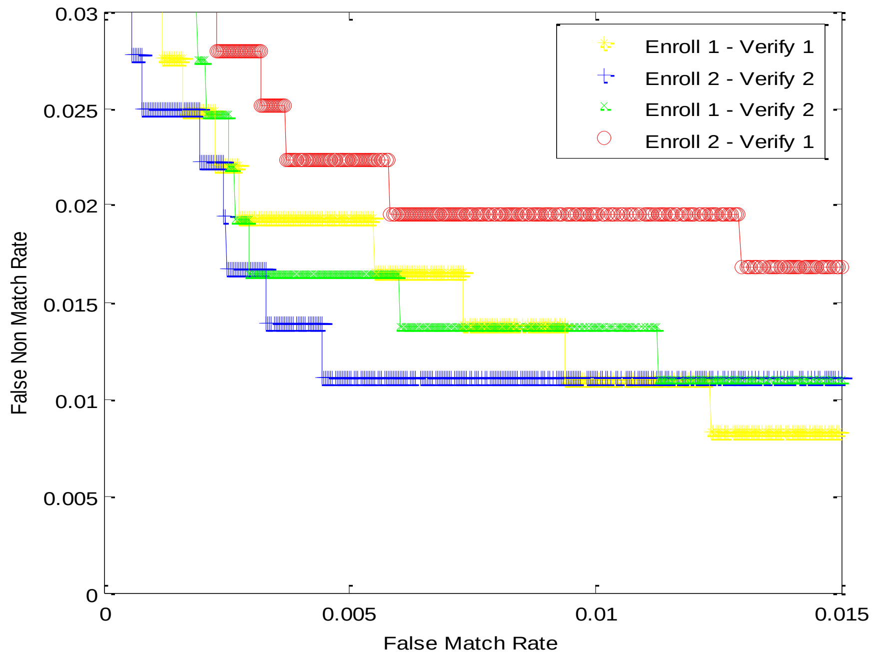


Figure 155. Enroll I Verify B – GFRR Versus GFAR

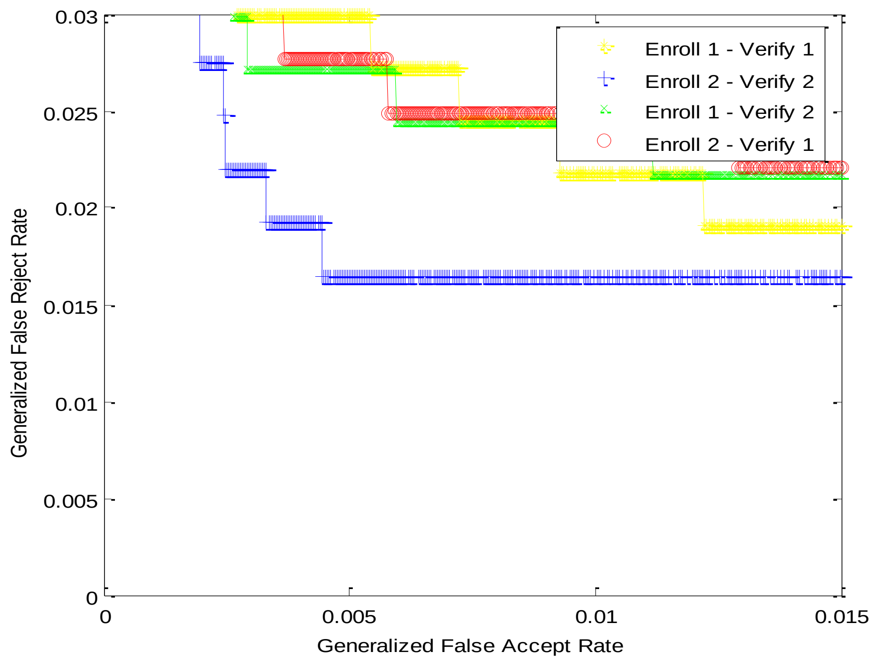


Figure 156. Enroll I Verify C – FNMR Versus FMR

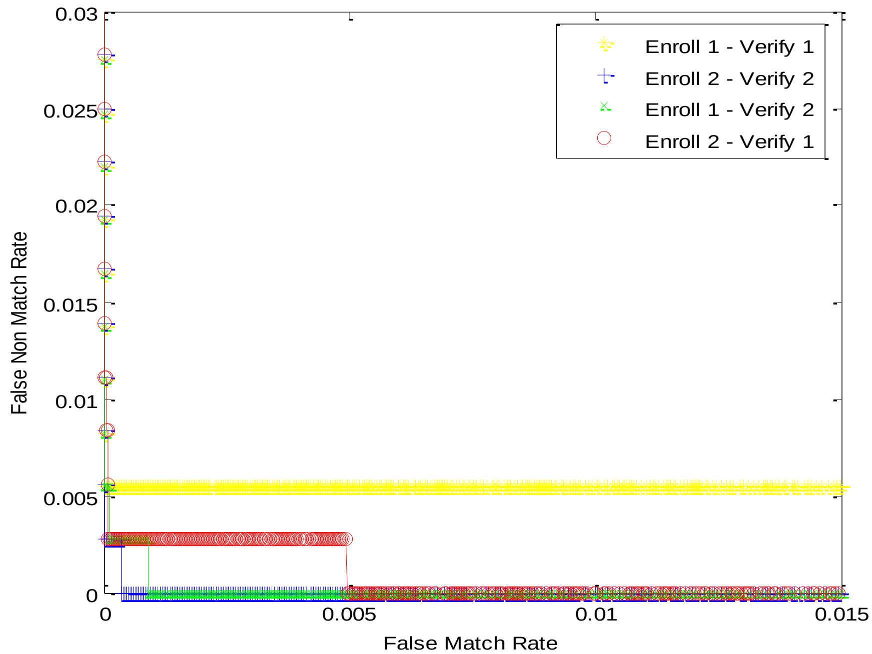


Figure 157. Enroll I Verify C – GFRR Versus GFAR

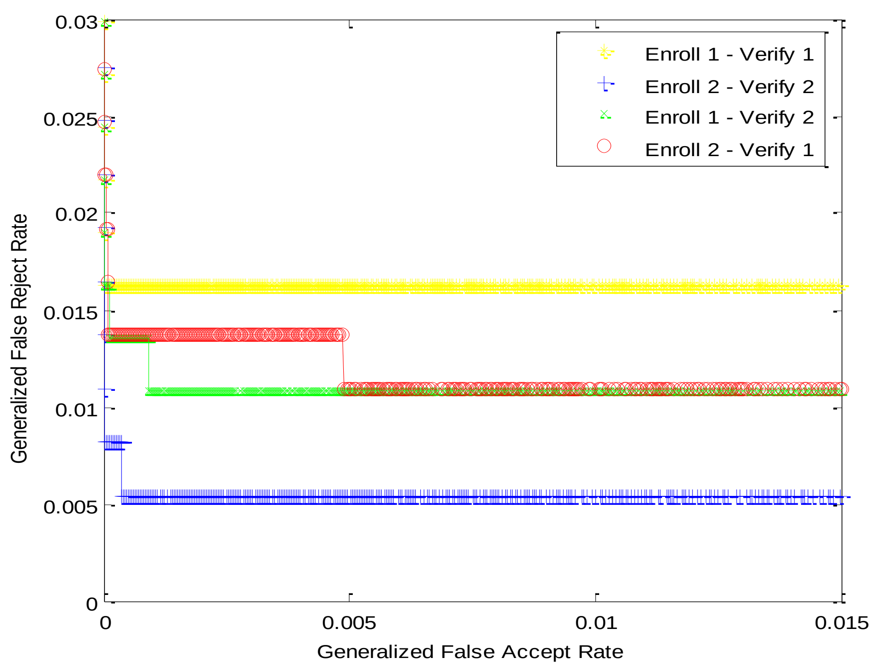


Figure 158. Enroll I Verify D – FNMR Versus FMR

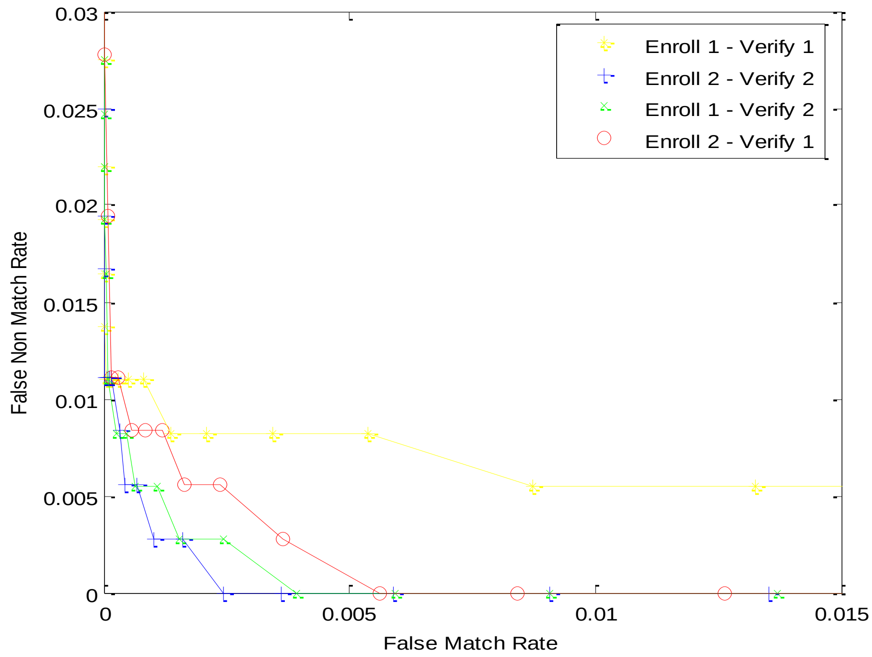


Figure 159. Enroll I Verify D – GFRR Versus GFAR

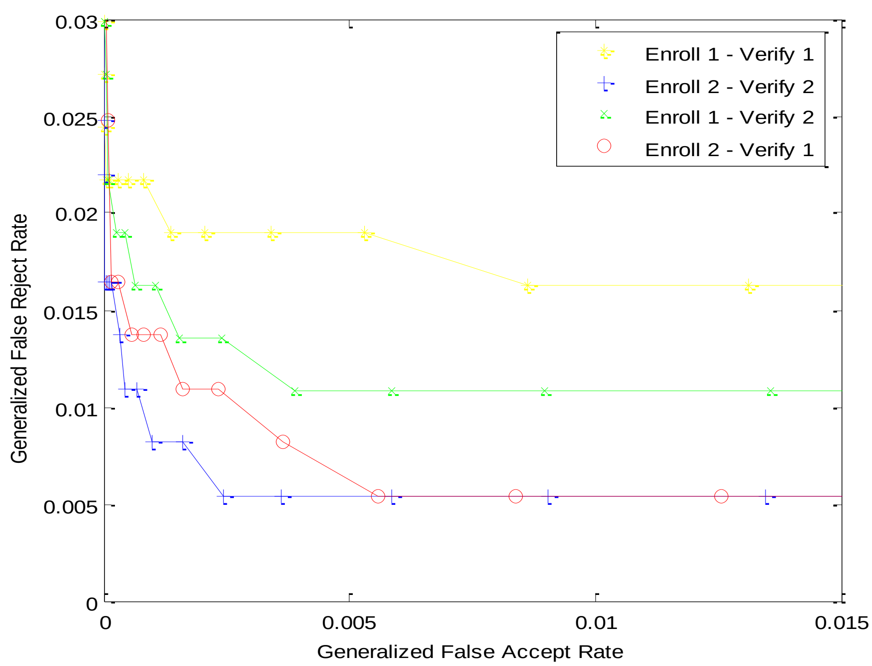


Figure 160. Enroll I Verify E - FNMR Versus FMR

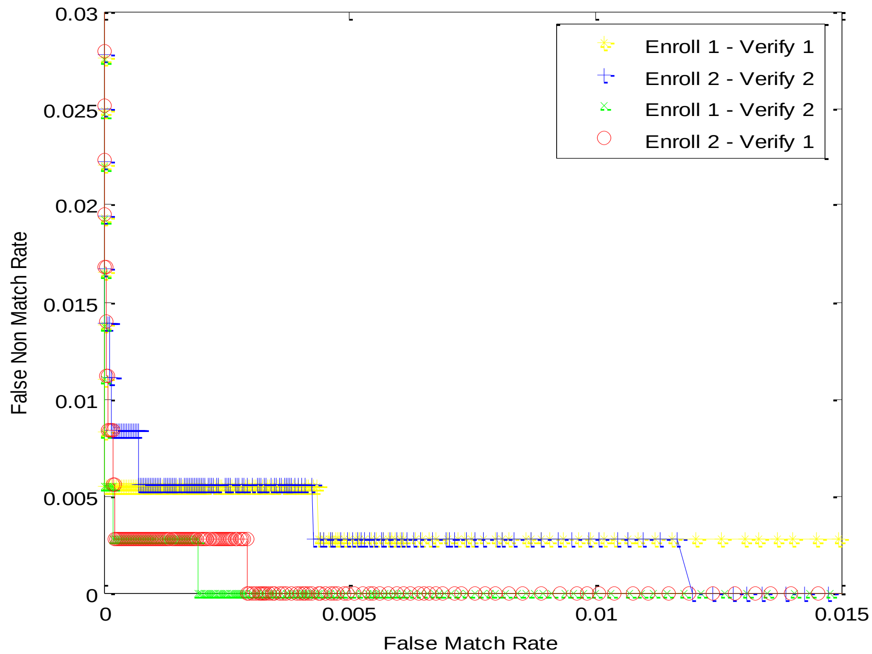


Figure 161. Enroll I Verify E – GFRR Versus FMR

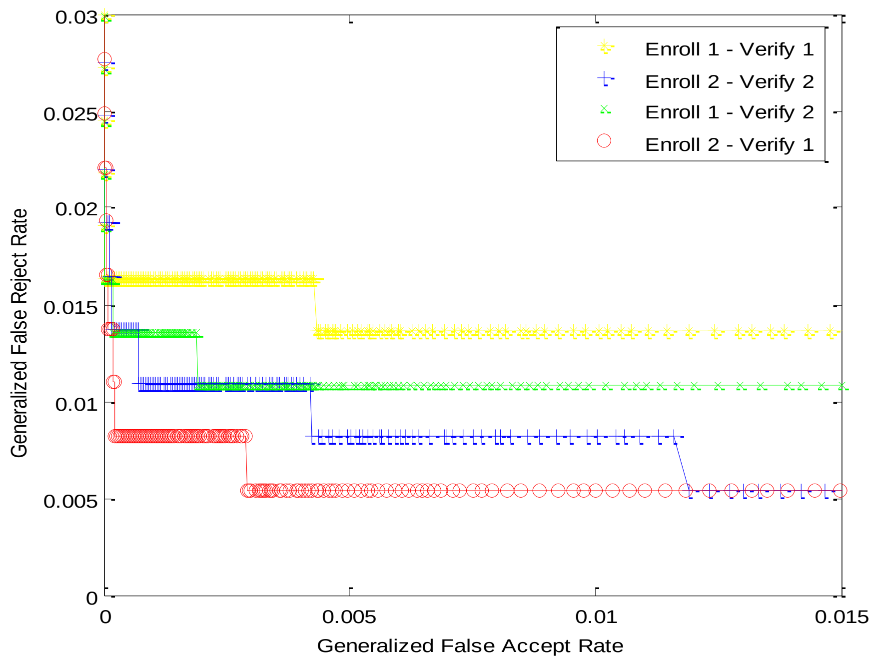


Figure 162. Enroll I Verify F – FNMR Versus FMR

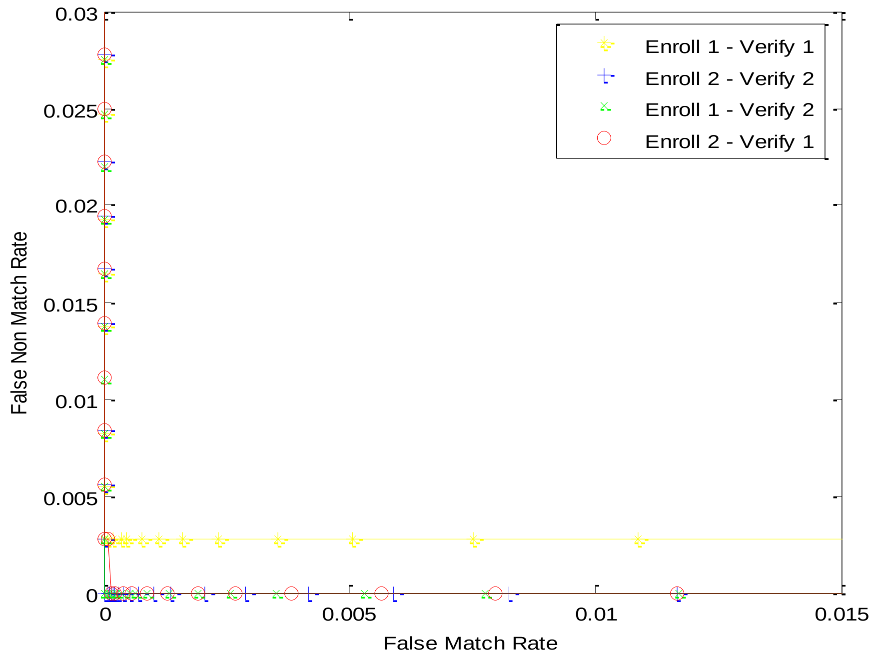


Figure 163. Enroll I Verify F– GFRR Versus GFAR

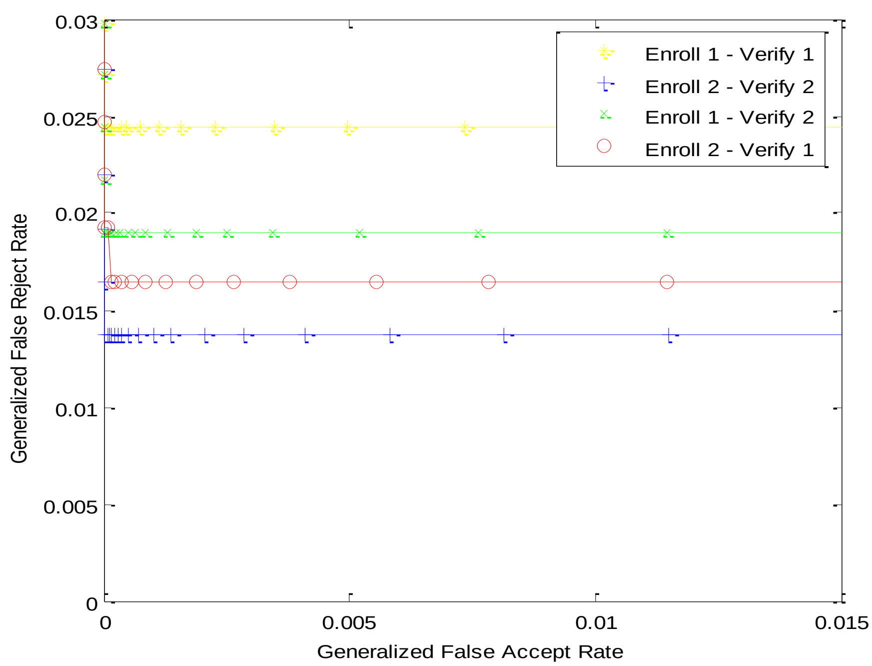


Figure 164. Enroll I Verify G– FNMR Versus FMR

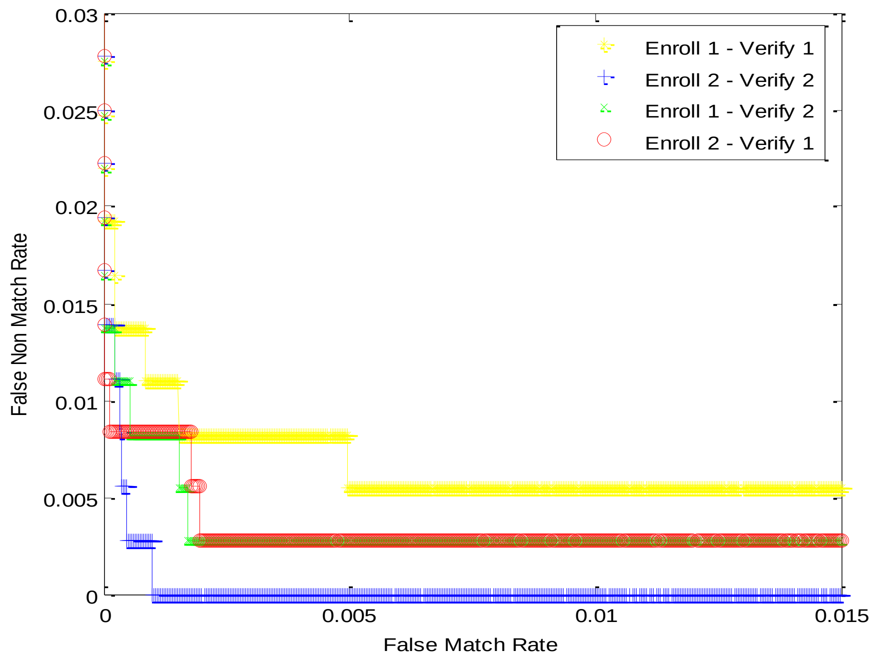


Figure 165. Enroll I Verify G – GFRR Versus GFAR

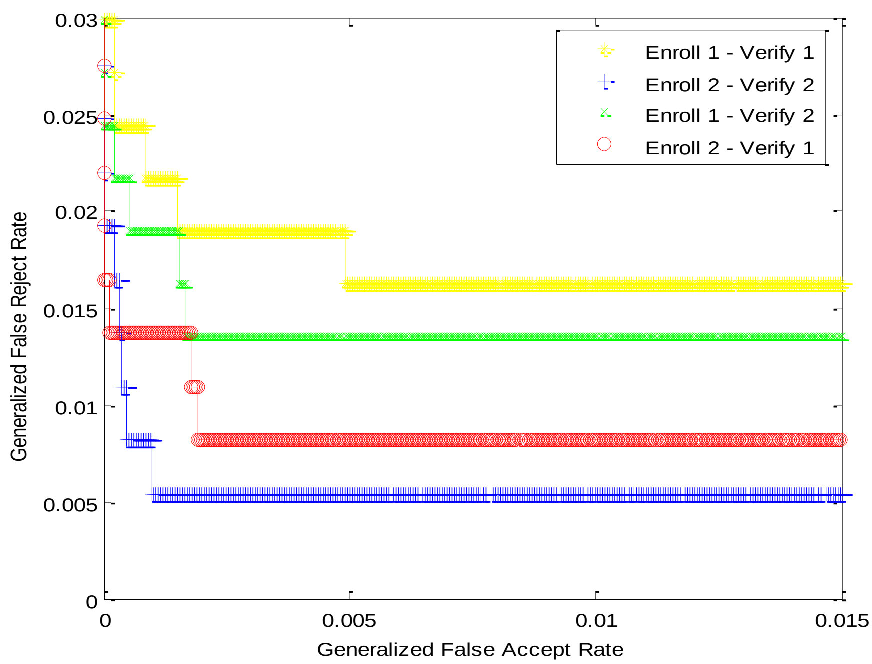


Figure 166. Enroll I Verify H – FNMR Versus FMR

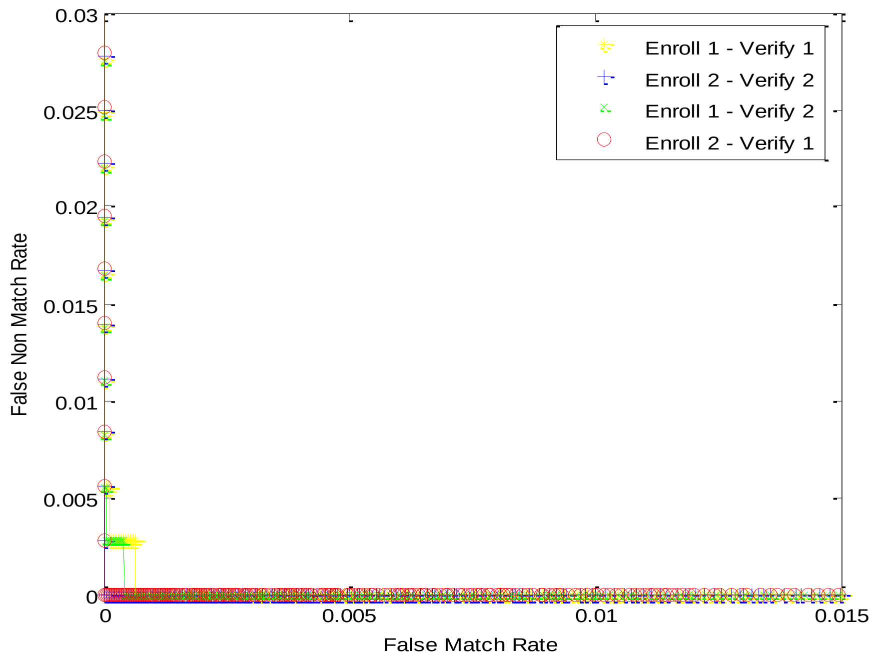


Figure 167. Enroll I Verify H – GFRR Versus GFAR

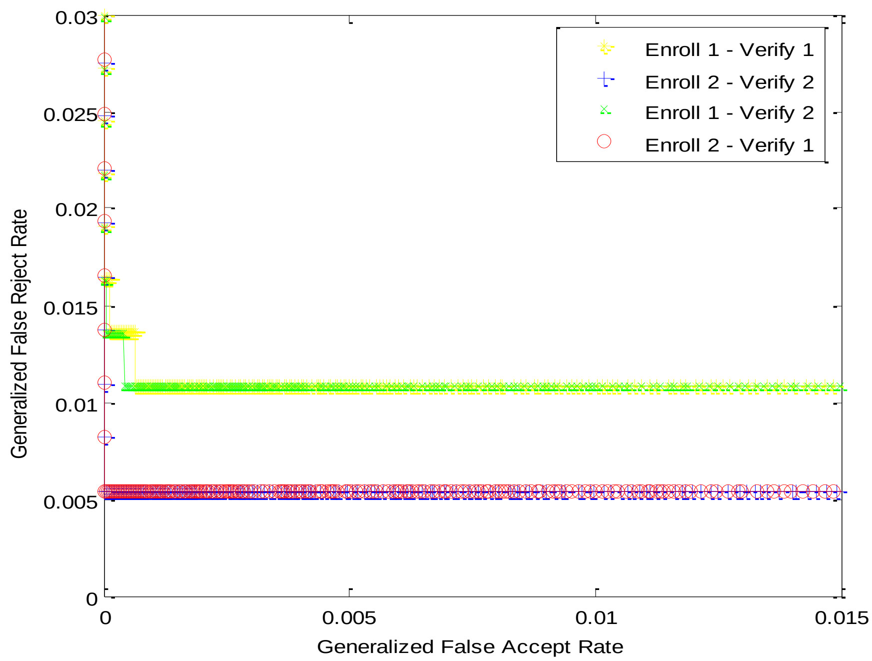


Figure 168. Enroll I Verify I – FNMR Versus FMR

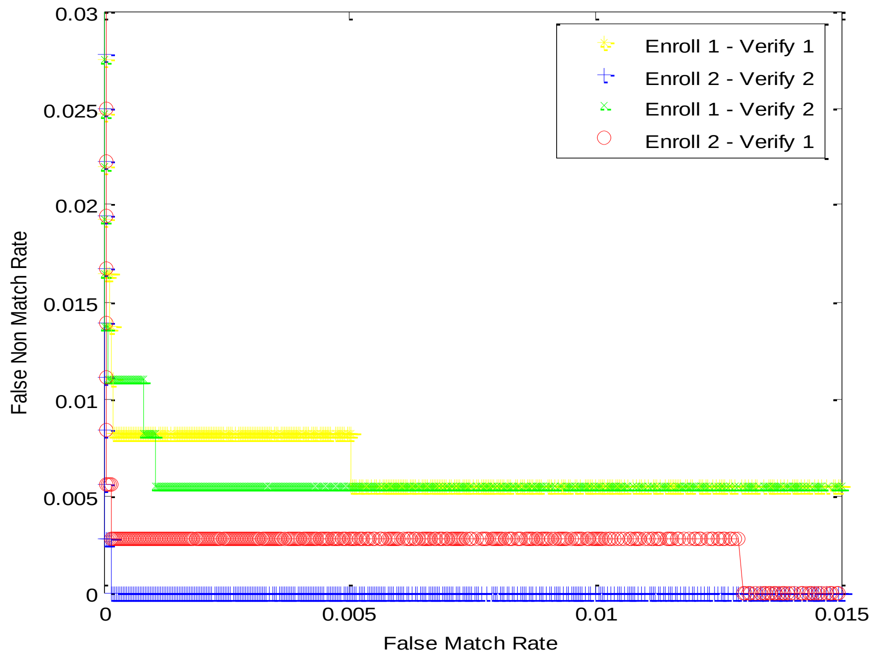


Figure 169. Enroll I Verify I – GFRR Versus GFAR