AAMVA UID9 Biometric Identification Report

Phase I: Technical Capability of Biometric Systems to Perform 1:300m Identification

Final Report

International·Biometric·Group
Research Consulting Integration

Version	Description	Revisions	Pages	Date
1.0	Draft delivered for review	n-a	n-a	7-18-03
2.0	Draft incorporating UID9 comments, questionnaire responses	*in toto*	*in toto*	8-01-03
Final Report	Incorporates comments from UID9 Denver meetings, additional revisions	*in toto*	*in toto*	11-12-03

Deleted: AAMVAUID9

Deleted: Draft 1.0

Table of Contents

Deleted: AAMVAUID9

Deleted: Draft 1.0

Deleted: AAMVAUID9

Deleted: Draft 1.0

Table of Figures

Deleted: AAMVAUID9

Deleted: Draft 1.0

1. Executive Summary

Project Background

International Biometric Group (IBG) was engaged by the American Association of Motor Vehicle Administrators (AAMVA) in April 2003 to assist the AAMVA Unique Identifier Task Group (UID9) in determining if biometric technology can be successfully used to perform identification against a biometric database comprised of 300m records. This number of records roughly corresponds to that of a fully populated database incorporating DL/ID records from all AAMVA jurisdictions. Identification, also referred to as "1:N", is the process by which a database is searched for multiple records belonging to the same individual. This Report is one of several documents developed in the course of this project designed to address the question of 1:300m identification.

Report Objective

Phase I of the multi-phase evaluation, presented in this Report, examines evidence relevant to determining whether biometric technology can be successfully used to perform 1:300m identification. Based on this evidence, this Report attempts to determine whether one or more biometric technologies – or a combination of biometric technologies – can be successfully used to perform 1:300m matching. Phase I represents an evaluation of biometric technologies from a theoretical perspective: operational and environmental constraints present in DL/ID usage environments are out of scope.

1:300m Concept of Operations

The 1:300m application addressed in this Report involves collection of biometric data from an individual in the course of his or her interaction with a DMV employee. The biometric system is designed to operate such that biometric data acquired from customers who have not previously enrolled should not match against any records in the database. If the 1:N process does not result in a match, the issuing agency can be confident that the customer has not already established a biometric identity. Customers who *have* previously enrolled, but are attempting to fraudulently enroll their biometric data again (most likely with a different set of personal information), are expected to match against their previous enrollment. The size of the database being searched, as well as the number of records being enrolled in the database over a given period of time, as central parameters that define

Deleted: AAMVAUID9

Deleted: Draft 1.0

the scale of this effort. Estimates developed external to this document[1] utilize an estimated peak transaction rate of 130,000 enrollees per day across North American jurisdictions.

Biometric Technologies Evaluated

This Report addresses the three biometric technologies with histories of deployment in reasonably large-scale identification applications: **facial recognition, fingerprint, and iris recognition**. While other biometric technologies have been deployed for identification on modest databases, notably speaker verification, these three technologies are commercially available as 1:N technologies and have been deployed in 1:N implementations with over 100,000 enrollees. An introduction to biometrics and biometrics processes is provided in Appendix B.

Two types of fingerprint systems based on the same underlying matching technology are addressed in this document: rolled and flat fingerprints. Rolled fingerprints are images of the entire fingerprint from “nail to nail”, acquired through a time-consuming, labor-intensive process. Rolled fingerprints are used in background checks and criminal booking processes. Flat fingerprints, used in civil ID applications, are images of the central area of the fingerprint. ”Flats” are acquired through a faster and simpler process than rolled fingerprints, and do not require manual handling of enrollee fingers (a requirement in rolled fingerprint systems). Rolled fingerprint images provide a larger quantity of biometric data and thus are capable of higher accuracy than flat fingerprints, although the exact difference in accuracy has not been extensively studied. Fingerprint systems can also vary in number of fingerprints acquired, which fingers are acquired, method of acquisition, and placements per finger.

1:N facial recognition and iris recognition systems are implemented in a relatively consistent fashion. Facial recognition systems generate candidate lists and often use a single facial image for enrollment and matching. Iris recognition systems generally provide much more accurate identification, using several images of the iris. Combinations of different biometric technologies known as **multimodal solutions**, may provide sufficient accuracy to enable accurate 1:300m identification, but few of these solutions have been deployed or tested.

Evaluation Methodology

Data presented and analyzed in this Report is gathered from biometric tests, deployments, and technology providers. Results from 1:N biometric **test efforts** provide a strong indicator

[1] **Proposed Measurement Ranges, Draft 0.4, Fischer Consulting Inc.**

Deleted: 3
Deleted: AAMVAUID9
Deleted: Draft 1.0

of a technology's ability to perform 1:N identification. Results must be evaluated in the context of a test's methodology in order to determine their applicability to real-world performance against large-scale databases. 1:N deployments are evaluated to assess the frequency with which technologies are utilized to perform large-scale identification in operational environments, as well as deployments' success rates. Deployment information is collected through public announcements, publicly disclosed projects, and other public sources. Information from **technology providers**, including developers, vendors, and integrators of technologies used in 1:N identification, provide insight into the long-term viability of performing very large-scale identification.

Performance Evaluation Parameters

1:N performance is traditionally measured in biometric systems according to **false match rate (FMR), false non-match rate (FNMR),** and **failure to enroll rates (FTER)**. FMR and FNMR are imperfect measurements of 1:N performance, as they do not account for all possible 1:N search outcomes.

In identification systems, "effective FMR" refers to the projected frequency with which an enrollee will erroneously match against one or more previous enrollees. False matches can inconvenience customers and require additional investigative resources. "Effective FNMR" refers to the percentage of individuals capable of successfully establishing a second identity. FTER represents the rate with which an individual will be unable to enroll in a biometric system, for example due to damaged fingerprints. Biometric systems are designed such that adjusting security thresholds to reduce a particular error rate will increase other error rates; administrators must manage this performance tradeoff in real-world systems.

An alternative method of evaluating identification system performance, referred to as **Open Set Identification Performance Metrics**, more effectively categorizes 1:N performance. The five metrics are (1) *Individual In Database Identified*, (2) *Individual In Database Identified As Another Individual*, (3) *Individual In Database Not Identified*, (4) *Individual Not In Database Identified As Being In Database*, and (5) *Individual Not In Database Not Found In Database*. Unfortunately these metrics are new to the biometric industry, and almost no performance data has been published in this fashion. UID9 sees use of these metrics as critical to understanding technologies' ability to address its requirements.

Deleted: AAMVAUID9

Deleted: Draft 1.0

Biometric Tests Types: Technology, Scenario, and Operational

No widely accepted standards for executing biometric tests or reporting biometric results exist. Therefore results from one test may be difficult to compare to another. The three generally accepted types of biometric tests are *technology*, *scenario*, and *operational.*

Technology testing involves comparison of databases of biometric records by means of one or more matching algorithms. By using static data, technology tests are repeatable, and do not require that the test subject be present. Technology testing is highly applicable to 1:300m system evaluation: large biometric databases of can be used to test multiple algorithms and determine how accurately technologies can scale. Technology testing utilizes probe and gallery databases – in a DMV application, the probe corresponds to new enrollees, while the gallery corresponds to the database of enrollees.

Scenario testing involves collection and matching of biometric data from test subjects in a controlled test environment, emulating a specific biometric usage scenario. Scenario testing evaluates full biometric systems as opposed to only algorithms. By virtue of its use of small test population, scenario testing is not applicable to 1:300m system evaluation. **Operational testing**, the least common of the three types, measures biometric performance in an actual deployment with real users. Variables that impact biometric performance are incorporated within operational tests results. The lack of control over the user population greatly limits the ability to effectively measure operational tests performance.

Synopsis of Test Results for Facial Recognition, Fingerprint, and Iris Recognition

More extensive test results have been published for facial recognition than any other biometric technology. Tests results strongly suggest that facial recognition cannot successfully perform single-record identification on a database of 300m users, and that conducting facial recognition searches on a database of 300m with a typical daily enrollment load would result in a large number of matching errors.

In spite of fingerprint technology's broad deployment, few independent tests of fingerprint identification technologies have been published. There is insufficient data to draw any conclusions about the scalability of fingerprint technology from published test results.

Less independent testing has been conducted on iris recognition than for facial recognition or fingerprint. The limited testing conducted suggests that iris recognition may be capable of accuracy on the order required for 1:300m matching, but insufficient data is available to

Deleted: AAMVAUID9

Deleted: Draft 1.0

support this thesis. Methodologies and results for fingerprint, iris, and face tests are synopsized as follows.

Biometric	Methodology Synopsis	Results Synopsis
Facial Recognition	Highly structured technology tests executed, providing results for different vendors	Poor 1:N results on modest databases cast substantial doubt on 1:300m performance
Fingerprint	Highly divergent test methodologies not highly applicable to 1:300m identification	Little 1:N test data published, considering breadth of deployment; insufficient test data to draw conclusions
Iris Recognition	Vendor-executed technology tests provide useful data, but lack of FNMR and FTE data limits applicability	Results highly promising, even for a single iris; incomplete results in terms of FNMR (as well as lack of third party testing) are limiting factors

Figure 1: Test Methodology and Results Synopsis

Biometric Tests: Facial Recognition Vendor Test (FRVT) 2002

Test Overview. FRVT 2002 tested 8 state-of-the-art facial recognition systems in mid-to-late 2002. From UID9's perspective, a key FRVT 2002 test was to evaluate systems' ability to return the correct record as the first match in a 1:N search. However, all tests involved subjects already enrolled in the test database: this differs from the 1:300m concept of operations in which most new enrollees as not already enrolled in the database. The test did not measure FTER those unable to enroll were classified as false non-matches.

Test Parameters. FRVT 2002 used 121,589 images from 37,437 individuals. Images were taken from Department of State visa applications, and were acquired in a controlled, operational environment. Nearly all enrollees were of Mexican decent. The time lapse between facial images used for testing ranged from under 60 days to nearly 3 years. Results were categorized by age and gender.

Test Results. FRVT 2002 demonstrates the difficulty of 1:N searches on even modest databases.

- The best system identified 26% of multiple enrollees as the wrong individual, such that the true multiple enrollee would have been undetected.
- Performance trends against increasingly large databases suggested that 60% of multiple enrollees would be misidentified in a database with 10m-100m enrollees.
- All systems had lower FNMR for males than for females.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

- Most systems recognized individual aged 50 years and over better than they did younger people; error differentials ranged from a few percentage points to 4 times higher.
- For the best systems, the correct ID rate dropped from over 80% within 120 days of enrollment to <70% after 1 year and <60% after 3 years.

Formatted: Bullets and Numbering

Results from FRVT 2002 should be reasonably applicable to DL/ID environments, given the manner in which facial images was captured. Results strongly suggest that facial recognition cannot reliably perform identification against databases of hundreds of millions of individuals. Strong trends within demographic groups suggest that certain groups of individuals would be matched much more reliably than others. Furthermore, the FRVT 2002 test methodology does not map directly to UID9 requirements.

Formatted: Bullets and Numbering

Biometric Tests: Face Recognition at a Chokepoint

Test Overview. Face Recognition at a Chokepoint was a Department of Defense scenario test that evaluates a single system's ability to perform watchlist identification, comparing individuals in real time against a gallery. This test is relevant to the 1:300m effort inasmuch as it is the only test to incorporate the type of performance measurements needed in a 1:N application. However the scenario test uses a very small population and is based on video as opposed to static images.

Test Parameters. The test utilized databases of 100-1,575 users, against which 144 individuals attempted to match. Approximately ½ of the 144 were present in the databases being searched. Test subjects stop and look at test cameras for approximately three seconds, then continue walking. The system continuously compares faces watchlist databases comprised of moderate- and high-quality facial images.

Test Results. Select test results indicate that on a 100-person database, a leading facial recognition system set at medium security performs as follows:

Formatted: Bullets and Numbering

- 53% of multiple enrollees were identified correctly
- 47% of multiple enrollees were not identified or were identified incorrectly
- 3% of first-time enrollees were incorrectly matched against an individual in the database

The comparatively small size of this database, and the error rates encountered, call into question the scalability of facial recognition for much larger systems.

Deleted: AAMVAUID9
Deleted: Draft 1.0

Biometric Tests: NIST Standards (Fingerprint)

Test Overview. NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability tested the ability of a generic fingerprint algorithm to identify correct records from a database of 600,000 enrollees.

Test Parameters. Testing took place against two fingerprint databases, each with over 600,000 subjects. Each database housed at least two left and two right flat index fingerprints from each enrollee. Images were acquired in border patrol environments and in Mexican Consular offices. 1000 subjects were chosen at random from the database to match against the 600,000-subejct database. Left and right index fingerprints were tested and reported separately. Testing utilized an open source fingerprint algorithm, such that results may not be indicative of vendor performance.

Test Results. Key results of the NIST Test were as follows.

- Database 1 (600k records): between 66% and 86% of were individuals correctly identified as first match; right index performed approximately 3-4% better than left index
- Database 2 (600k records): between 67% and 89% of individuals correctly identified as first match; right index performed approximately 10%-12% better than left index.

Formatted: Bullets and Numbering

Results demonstrate the need for multiple fingerprints to execute large-scale identification reliably. A single fingerprint is unable to identify the correct user between 12% and 34% of the time on a database of 600,000 individuals; real-world fingerprint systems designed for databases larger than 1m enrollees utilize multiple fingerprints. However, multi-fingerprint tests are rarely executed. NIST's recently announced Fingerprint Vendor Technology Evaluation[2] (FpVTE) will test multiple rolled and flat fingerprint images in late 2003. 50,000 tenprint records will be used to generate 2.5 billion matches. Test results may provide answers to critical performance questions and inform UID9 decisions.

Biometric Tests: Philippines AFIS Benchmark Testing

Test Overview. Philippines AFIS Benchmark Testing was conducted in the late 1990's to evaluate the ability of AFIS vendors to execute 1:N fingerprint classification and matching. This test projects performance of multiple fingerprint systems relevant to large-scale database requirements.

Test Parameters. Roughly 4000 fingerprints were taken from over 500 individuals under

[2] http://fpvte.nist.gov/index.html

Deleted: AAMVAUID9

Deleted: Draft 1.0

highly controlled and directed conditions. Fingerprints were cross-compared to generate approximately 16m matches. The four participating AFIS vendors classified records according to fingerprint characteristics to limit the number of records to be searched. Vendors then cross-compared fingerprint databases to generate match scores.

Test Results. The best single-finger systems showed zero false matches with an FNMR of less than 10%. Based on the overall data available, test authors project that a single-finger FMR of 1/1,000,000, with a corresponding FNMR of 10%, is attainable by the strongest performers. Authors also project that a two-fingerprint system can address deployer requirements for scalability to over 8m records. No statement is made regarding matching on databases of 100m records.

Biometric Tests: Iridian Cross-Comparison Test

Test Overview. The Iridian Cross-Comparison Test represents tests the susceptibility of iris recognition to false matching in 1:N applications. This testing is relevant inasmuch as it is the first large-scale test of iris recognition technology ever published.

Test Parameters. Approximately 110,000 iris templates were collected from operational databases in the U.S, Middle East, and South Asia. A separate 9000-person database was compared against these records, such that nearly 1b matches were executed. The primary results reported were the number of matches at specific security levels (referred to as Hamming Distances, or HDs. Smaller HDs indicate more similarity between templates and thus higher security levels. Templates with no similarity would have a HD of 0.50.

Test Results. Results indicate the number of false matches at different security thresholds.

- **157** of the 1b comparisons resulted in an HD at or below 0.31
- **32** of the 1b comparisons resulted in an HD at or below 0.30
- **10** of the 1b comparisons resulted in an HD at or below 0.29
- **3** of the 1b comparisons resulted in an HD at or below 0.28
- **1** of the 1b comparisons resulted in an HD at or below 0.27
- None of the 1b comparisons resulted in an HD less than 0.26

By enforcing certain decision policies in large database searches, Iridian projects an effective FMR of less than 1 in 2.79m for searches against a database with 100m records. Such performance is well within the bounds established for FMR in discussions with UID9. Furthermore, this test utilized only one iris from each individual. Two irises may provide a near-multiplicative effect on accuracy. A major gap in test results is the lack of FNMR data.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

In order for a test's FMR to be meaningful, it is necessary to also measure FNMR to assess the percentage of multiple enrollees who would have evaded detection. The testing also did not measure FTE: all individuals tested were by definition able to enroll.

Deployments

Deployment characteristics and applicability to 1:300m matching are synopsized as follows:

Technology	Deployment Characteristics	Deployment Applicability
Facial Recognition	Small number of implementations over 1m; deployments generate candidate list, not single records	Concept of operations applicable to DMV environment, though candidate list generation a drawback
Fingerprint	Several implementations over 10m; rolled systems deployed more often than flat, though a 60m enrollee system is based on flat prints; capable of single-record location	Highly applicable to 1:300m application due to field scalability and single-record location concept of operations
Iris Recognition	Deployments with between 100,000 and 500,000 enrollees; largest implementations in Middle East, South Asia	No deployments with over 1m records; however, concept of operations is consistent with UI(requirements

Figure 2: Deployment Characteristic and Applicability Synopsis

No current biometric deployments even begin to approach the scale of the 300m record system under consideration. The largest fingerprint deployments are in the tens of millions; the largest facial recognition deployments are in the millions; and the largest iris recognition deployments are in the hundreds of thousands. Furthermore, performance data such as false match, false non-match, and failure to enroll rates from operational implementations is rarely available. This lack of available data may be attributable to inability to measure operational performance, confidentiality agreements with technology providers, or risks of disclosing performance gaps that could enable system circumvention.

Deployments: Facial Recognition

Facial recognition's largest deployments are in DL/ID and passport issuance applications. These deployments are based on conversion of existing image databases to searchable biometric records. In these applications, large databases are searched to generate candidate lists of potential matches; potential matches are resolved manually. Among the largest 1:N facial recognition implementations are the following:

- Illinois DMV: 13m records *(eventual scale: 25m records)*

Formatted: Bullets and Numbering
Deleted: AAMVAUID9
Deleted: Draft 1.0

- Colorado DMV: 10m records
- Australia Passports: 3.5m records

No currently publicized facial recognition deployments approach the scale of the 300m application under consideration. However, a broad range of larger facial recognition deployments is likely to emerge with the adoption of biometrically-secured passport issuance processes. Between today and October 2004, several databases with over 10m records will likely be populated in the course of passport and visa issuance.

Deployments: Fingerprint

Many but not all of the largest fingerprint systems are criminal ID systems, based on acquisition of rolled fingerprint images, as opposed to civil ID systems that acquire flat placements. Among the largest 1:N fingerprint implementations are the following:

Formatted: Bullets and Numbering

- Nigeria National ID: 60m records (six fingerprints, flat placement)
- FBI Integrated AFIS (IAFIS): 55m records (ten fingerprints, rolled)
- Malaysia Government Card: 18m records (two thumbprints, flat)
- California Department of Justice: 12m records (ten fingerprints, rolled)

The largest of these systems in Nigeria was populated in 2003, such that performance data is not yet available. Although several fingerprints deployments exceed 10m records, even fingerprint technology – by far the most established biometric technology – does not yet approach the scale of the 1:300m application under consideration. Dedicated acquisition processes presents challenges not present in facial recognition systems, where existing acquisition processes can be leveraged. Fingerprint systems used in civil ID applications locate single records as opposed to candidate lists, and are more relevant to UID9.

Deployments: Iris Recognition

Iris recognition is less widely deployed in large-scale databases than other technologies capable of performing 1:N identification. Among the largest 1:N iris recognition implementations are the following:

Formatted: Bullets and Numbering

- United Arab Emirates: 200,000 records (eventual scale: 1m records)
- Pakistan System 2/1: 120,000/65,000 records (eventual scale: 1m records)

Iris recognition providers have populated internal databases comprised of over 100,000 records. However, even the largest operational iris recognition databases are less than 1/1000th the size required in a 300m record database. Larger databases are very likely to

Deleted: AAMVAUID9

Deleted: Draft 1.0

emerge: in early 2002, the largest iris recognition databases were in the low tens of thousands, demonstrating the increased adoption of the technology. Potential projects using iris recognition in national ID programs could quickly result in deployments of over 1m users.

Performance Data from Technology Providers

Vendor-supplied performance projections applicable to 1:300m applications can be synopsized as follows:

Technology	Performance Projections	Applicability
Facial Recognition	Estimated 69% correct ID rate on 300m database	Based on internal tests and third party evaluations; assumes linear decrease in FNMR
Fingerprint	Projected 99.5 correct ID rate on 300m person database, using 4-6 fingerprints	Based on internal testing; vendor states that methodology was developed to evaluate scale to 300m
Iris Recognition	Projected 98-99% correct ID rate on 300m person database, with less than 1 in 6m FMR; assumes 1 iris	Based on internal tests; use of both irises should substantially increase accuracy

Figure 3: Technology Provider-Supplied Performance Data Synopsis

Technology providers were surveyed regarding their ability to perform 1:300m identification. Certain fingerprint, iris, and facial recognition providers see their technology as being capable of addressing the problem of 1:300m identification. The following projections are representative of those provided by vendors who chose to project performance in a 1:300m application.

- **Facial Recognition:** estimates based on third party tests on 3.5m images, mapped to new algorithm capabilities, project identification rate of 69% against a 300m person database
- **Fingerprint:** estimates based on internal testing against a database of 100,000 records indicate that 4-6 flat fingerprints can be used to identify 99.5% of duplicate records in a 300m database (though acquisition of all ten fingerprints is recommended)
- **Iris Recognition:** estimates based on internal testing indicate that a single iris can provide an identification rate of 98-99%, while falsely matching approximately 1 in 6m enrollees, on a 300m person database

Some but not all vendors who refrained from providing performance projections in 1:300m applications cautioned against the use of extrapolated test data to project performance on larger databases. Other vendors noted the divergence in performance projected under

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

controlled conditions, in which one can optimize performance to suit the characteristics of a target database, as opposed to performance in operational databases where less control is available over the data.

Initial Evaluation of Technology Capabilities

Despite the lack of data indicative of 1:300m performance, provisional comments can be made regarding various technologies' capabilities in executing 1:300m matching. Results from small-scale and large-scale tests indicate that **facial recognition** will not be capable of successfully performing 1:300m identification. The technology is subject to errors in which a substantial percentage of multiple enrollees are not located in databases of hundreds or thousands of users.

Based on deployments of **fingerprint** systems, multiple flat-fingerprint solutions – particularly those which acquire 6, 8, or 10 fingerprint images from each enrollee – may be capable of identification on the order of 100m+ records. Multiple fingerprints increase the quantity of distinctive data associated with a given enrollee. Little independent data on the scalability of such solutions is available. We are unaware of any biometric system in operation based on more than 6 flat fingerprints; testing of such systems is just beginning.

Testing conducted by Iridian indicates that **iris recognition** decision policies can be enforced such effective FMR (the projected rate with which an individual falsely matches against a 1:N database) to less than one incident per million, much lower than UID9 guidelines are likely to recommend. The relevance of this data is uncertain due to the lack of corresponding false non-match data, without which the above false match data cannot be reasonably assessed.

Limitations of Available Information

The following are the primary limitations in information available to assess biometric capabilities against 300m-record databases.

Lack of applicability of biometric test data. While biometric testing remains the primary means by which one can gain controlled, objective test data on biometric system performance, it is difficult to draw definitive conclusions on performance in 1:300m applications for several reasons: (1) Only a small number of tests have been executed and published for each of the three primary 1:N technologies; (2) no testing of systems with millions or tens of millions of records has been executed; (3) most 1:N test methodologies generate results inconsistent with the concept of operations of DL/ID transactions, being

Deleted: AAMVAUID9

Deleted: Draft 1.0

primarily focused on multiple enrollee record location; and (4) most biometric testing fails to report results over time. In addition, population composition, acquisition and matching methods, system costs, deployer business and functional requirements, usage environment, and deployment/enrollment timeframe can have a direct and pronounced impact on performance difficult to measure in biometric tests.

Lack of reliable information gained through deployments. Balanced performance data is difficult to gather through real-world deployments, due to the difficulty of isolating and controlling variables in specific usage environments. False non-match rates, by definition, are difficult to assess in an operational system.

Lack of precedent in addressing systems of this scale. No vendor has ever had to address, in an actual deployment, the challenge of 1:300m identification. Vendors therefore may not have reliable or useful information on the degree to which their technology can scale to this level, aside from purely mathematical extrapolations.

Difficulty in providing abstract performance estimates. The hypothetical nature of the 1:300m question is such that unanticipated performance considerations may come into play once one reaches very large database sizes. Projecting performance from smaller operational or test databases may thus be inconsistent with real-world performance. Technology providers with a business focus on addressing issues involved in real-world systems may struggle to project performance in a theoretical system that lacks real-world constraints around matching and acquisition. Lastly, vendor data may prove to be excessively optimistic, ungrounded, non-reflective, or unrealistic performance estimates. In most cases no method will be at hand to qualify the usefulness of vendor claims.

Impact of Emergence of Large-Scale Civil ID Systems

The composition and prevalence of large-scale systems is expected to change dramatically over the next three years with the emergence of large-scale civil ID projects such as border entry/exit, national ID, and biometrically-enabled passports. Facial biometrics are to be incorporated in passports and machine-readable travel documents around the world, in accordance with ICAO recommendations and in compliance with U.S. legislative requirements. This will increase by orders of magnitude the number of individuals enrolled in facial recognition systems (although the processes by which 1:1 and 1:N matching will take place for such systems is as yet undetermined). Iris recognition is currently positioned for deployment in border control projects in the Middle East, in South Asia, the U.K., and

Deleted: AAMVAUID9
Deleted: Draft 1.0

Canada. It is very likely that systems with well over 1m iris recognition enrollees will be populated within this timeframe. Therefore any decisions informed by the current lack of large-scale deployments of facial recognition and iris recognition technologies may be superceded by future events. In addition, the adoption of multiple-fingerprint systems is expected to increase over the next three years, although the rate of adoption for fingerprint may not increase as rapidly as for facial and iris systems.

Focus Areas for Further Evaluation

The following areas warrant further investigation as methods of improving biometric system performance and addressing the 1:300m challenge.

A **systematic and structured approach** to evaluating biometrics in a fashion consistent with the application under consideration is necessary. The large variance in biometric test methodologies and results reporting methods limits the ability to compare and contrast published test results. Generating performance data relevant to UID9 requires that problems of test scale and design be addressed. At least the tens of thousands of test subjects must be involved to begin to project performance on a 300m person database. Designing tests to mirror DMV acquisition, instruction, and matching processes is equally important.

Performing 1:300mtesting through **standardized data formats, processes, and systems** so as to avoid defining a proprietary approach that favors one vendor is a necessity to determine the degree to which performance degrades (if any) when moving from closed to open biometric systems. The development of shared interoperability standards, which fostered the use of fingerprints in law enforcement applications, can be leveraged for this application as well. A particular area of interest is the performance of iris recognition in independent tests. Vendor-executed testing has demonstrated considerable promise in the ability to limit FMR on large-scale searches.

Developing methods of combining match results from different biometric technologies via **multimodal systems** warrants research due to the opportunity to combine fingerprint, face, and/or iris data to achieve desired accuracy for 1:300m systems. Also, understanding the degree to which **fusion of multiple algorithms** can improve performance may be essential to developing highly scalable and accurate systems. Fusion biometric systems are those which process a single biometric input through multiple processing and matching algorithms to generate an improved total result.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

2. Introduction

Formatted: Bullets and Numbering

2.1. Project Background

International Biometric Group (IBG) was engaged by the American Association of Motor Vehicle Administrators (AAMVA) in April 2003 to assist the AAMVA Unique Identifier Task Group (UID9) in determining if biometric technology can be successfully used to perform 1:N matching against a biometric database comprised of 300m records. This number of records roughly corresponds to that of a fully populated database incorporating DL/ID records from all AAMVA jurisdictions.

2.2. Project Structure: Phase I through Phase III

In accordance with UID9 requirements, the evaluation is structured as a multi-phased process. This Report is one of several designed to support the overall decision process.

Phase I: Technical Capability of Biometric Systems to Perform 1:300m Identification
Phase II: 1:300m Identification with Additional DL/ID-Oriented Technical Considerations
Phase III: Feasible Implementation Assessment

Phase I of the evaluation, presented in this Report, provides evidence relevant to determining whether biometric technology can be successfully used to perform 1:300m matching absent operational considerations specific to DMV environments. Based on this evidence, this Report attempts to determine whether one or more biometric technologies – or a combination of biometric technologies – can be successfully used to perform 1:300m matching. Phase I represents an evaluation of biometric technologies in a theoretical usage scenario, such that the operational constraints and conditions present in DL/ID usage environments are not considered.

The criteria used to define "successful" 1:300m matching have been developed through discussions with UID9. These criteria were originally seen as false match rates (FMR), false non-match rates (FNMR), and failure to enroll (FTE) rates, as well as measures such as history of large-scale deployment and long-term day-to-day use. However, in the course of Report development, it was determined that *Open Set Identification Performance Metrics*[3] provided a better assessment of biometric performance in UID9's identification environment.

[3] See Section 5.3 for a full discussion of this concept..

Deleted: AAMVAUID9
Deleted: Draft 1.0

If, based on evaluation of data presented in this Report, one or more biometric technologies or technology combinations is determined to be capable of addressing UID9 performance requirements, the evaluation proceeds to Phase II (see below). If it is determined that no biometric technology or combination of technologies addresses UID9 performance requirements, Phase II will be expanded to address biometric technology capabilities when using "filtered" 1:N searches, those that incorporate factors such as gender, age, and geographical location to facilitate 1:N searches. Assessing technologies along with filtering may or may not reveal biometric technologies or technology combination capable of addressing UID9 requirements.

Phase II of the evaluation places the 1:300m technology assessment in the context of an operational DL/ID system, incorporating considerations such as response time, resistance to influence by non-cooperative users, durability of acquisition devices, and ability to manage DL/ID enrollment loads. **Phase III** of the evaluation incorporates several non-technical factors, such as impact on current processes, costs, privacy impact, and deployment alternatives, in determining if a technology can be successfully deployed for 1:300m matching. These factors are broadly classified as those which support "feasible implementation."

2.3. Report Objectives

The objectives of the Phase I Report are as follows:

- To organize, present, and evaluate evidence available relevant to assessment of the viability of successfully performing biometric identification on a database of 300m records.
- If possible, to identify those technologies or technology combinations capable of successfully performing biometric identification on a database of 300m records.
- If necessary, to determine what further evidence must be collected or generated to assess the viability of successfully performing biometric identification on a database of 300m records.

2.4. Assumptions

The following assumptions frame the Phase I Report.

As the project commenced, the scale of the identification task was fixed at 300m records. This figure represents a general estimate of the number of DL/ID holders that would be

Deleted: <#>Phase I Report Organization ¶
<#>Report Objectives¶
<#>Assumptions ¶
<#>Evaluating 1:300m Capabilities: Methodology ¶
<#>1:300m Matching Concept of Operations ¶
<#>Approach to Data Collection and Analysis ¶
<#>Challenges and Limitations ¶
<#>Defining Evaluation Parameters ¶
<#>Error Types ¶
<#> Interrelation of Error Types ¶
<#>Biometric Technologies Capable of 1:N Matching ¶
<#>Facial Recognition¶
<#>Fingerprint ¶
<#>Iris Recognition ¶
<#>Multimodal Solutions ¶
<#>Test Efforts ¶
<#>Types of Biometric Testing ¶
<#>Facial Recognition Testing ¶
<#>Iris Recognition Testing¶
<#>Fingerprint Testing ¶
<#>Deployments ¶
<#>Overview ¶
<#>Facial Recognition ¶
<#>Fingerprint ¶
<#>Iris Recognition ¶
<#>Performance Data from Technology Providers ¶
<#>Facial Recognition ¶
<#>Fingerprint¶
<#>Iris Recognition ¶
<#>Conclusions and Recommendations: Assessing the Viability of 1:300m Identification¶
<#>Gaps in Understanding ¶
<#>Evidence at Hand to Facilitate Decisioning ¶
<#>Initial Evaluation of Technology Capabilities ¶
<#>Areas for Further Research ¶
<#>Appendix A: Large-Scale Testing Rollup ¶
<#>Appendix B: Fundamental Biometric Concepts and Processes ¶
<#>Appendix C : Biometric Identification Capabilities Questionnaire ¶

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

incorporated within a large-scale biometric system across AAMVA jurisdictions.

The range of potential technology solutions to the problem of 1:300m identification was expanded to include multimodal biometric solutions, those which utilize more than one biometric technology (e.g. fingerprint and face, or face and iris) for identification.

The Phase I Report is designed to address theoretical performance, absent the logistical and operational challenges involved in acquiring and processing data from such a large number of subjects. Therefore the evaluation of 1:N matching capabilities assumes an ideal biometric acquisition environment and optimal acquisition processes.

The Report assumes that 1:N matching will take place against newly collected data as opposed to legacy data (e.g. databases of existing photographs). This is consistent with the assumption that data will be collected in an ideal usage environment through optimal acquisition processes.

Deleted: In order to ensure an effective use of resources, this Report addresses the three biometric technologies commonly understood to be capable of performing large-scale identification: facial recognition, fingerprint, and iris recognition. While other biometric technologies have been deployed for limited-scale identification, notably voice verification, these three technologies are commercially available and have been deployed in 1:N implementations with over 100,000 enrollees. ¶
T

Deleted: AAMVAUID9

Deleted: Draft 1.0

3. Evaluating 1:300m Capabilities: Methodology

3.1. 1:300m Matching Concept of Operations

This Report is focused on the task of performing biometric identification against a database of 300m enrollees. Although most constraints and challenges involved in using biometrics in a DMV environment are to be addressed in subsequent Phases, certain elements of the 1:300m concept of operations must be defined in order to generate reasonable performance estimates and to provide meaningful evidence on the ability of biometrics to successfully execute 1:300m identification. For example, 1:300m covert identification of non-cooperative subjects is a completely different biometric task than cooperative identification of subjects deliberately interacting with a biometric system.

The 1:300m application addressed in this Report involves collection of biometric data from an individual ("customer") in the course of his or her interaction with a DMV agent ("operator"). The customer provides biometric data in a cooperative fashion under the direct or indirect supervision of an operator. The operator provides the necessary instructions to the customer to ensure effective data collection. The customer's biometric data may be evaluated in real time in order to determine if the quality is sufficient to generate a template. Insufficiently distinctive or measurable data requires that the customer provide data once again, although after a predetermined number of enrollment attempts a user may be declared a "failure to enroll" on the biometric system.

This reference set of biometric data is collected from the customer in order to execute a 1:N search against all previous enrollees. An enrollee is defined as a customer who has successfully provided such a biometric reference set. The biometric system is designed to operate such that customers who have not previously enrolled should not match against any records in the database, as their data will not be present. If no match occurs the issuing agency can be confident that the customer has not already established a biometric identity. Customers who have previously enrolled, but are attempting to fraudulently enroll their biometric data again (normally associated with a different set of personal information), should match against their previous enrollment. In most cases investigation would ensue, as the issuing agent has been alerted to the fact that two biometric records are similar enough to have been declared a match.

Deleted: Many of
Deleted: the specific
Deleted: not addressed
Deleted: until
Deleted: . However,
Deleted: report
Deleted: am
Deleted: AAMVAUID9
Deleted: Draft 1.0

Under this concept of operations, there is no scenario in which a previously enrolled user would be searched in a 1:N fashion unless he or she were attempting to enroll fraudulently[4]. That is, customers who claim identities, and whose biometric records exist in the database, are matched in a 1:1 as opposed to 1:N fashion. This is an important consideration, as biometric system configuration can differ depending on whether false matches or false non-matches need to be minimized. With the exception of clerical errors, all 1:N matches should be indicative of fraudulent enrollment attempts.

One last factor to be addressed in the concept of operations is match loads. The eventual size of the database being searched (300m) represents half of the equation when evaluating ability to successfully execute 1:300m identification. The second half of the equation is the number of records searched against this database over any given period (for simplicity, searches per day is a useful metric). Estimates developed external to this document[5] project a peak transaction rate 130,000 enrollees per day.

Deleted: a maximum yearly enrollment load of 32.5m, or

Deleted: maximum daily average of

Deleted: 3

3.2. Approach to Data Collection and Analysis

IBG has defined a tripartite data collection and analysis approach to the Phase I evaluation of biometric technologies' ability to successfully to perform 1:N matching in a system of 300m records. The three types of data collected are as follows:

- **Test-derived performance data**
- **Operational performance data**
- **Performance data projected by technology providers (i.e. vendors, integrators, developers)**

Information in these three areas is collected for each biometric technology and technology combination warranting consideration in a large-scale 1:N environment: facial recognition, fingerprint, and iris recognition (see ***Biometric Technologies Capable of 1:N Matching*** below for discussion of the selection of these technologies). This approach ensures the collection of a wide range of information.

[4] Barring clerical error

[5] **Proposed Measurement Ranges, Draft 0.4, Fischer Consulting Inc.**

Deleted: Proposed Measurement Ranges, Draft 0.3

Deleted: AAMVAUID9

Deleted: Draft 1.0

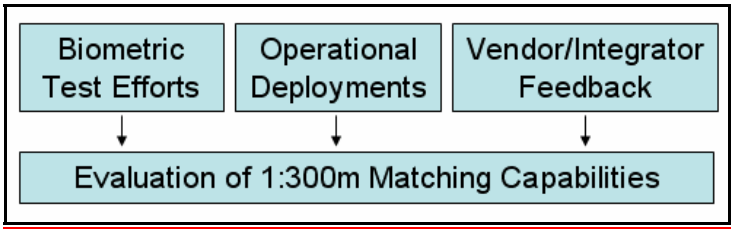


Figure 4: Data Collection Methodology

Deleted: ¶

3.2.1. Identify and Evaluate Relevant Biometric Tests

Results from 1:N biometric test efforts provide an indicator of a technology's ability to perform 1:N identification. By establishing controls on the manner in which biometric data is collected and compared, biometric tests can reduce the uncertainty associated with results from operational systems and can substantiate or disprove vendor claims as to a technology's scalability or performance. A number of biometric tests, including technology tests, scenario tests, and operational tests, have been executed over the past several years whose results may be applicable to the 1:300m identification question.

Test results must be evaluated in the context of a test's methodology in order to determine the degree to which they are likely to reflect real-world performance. This Report strives to incorporate results from biometric tests whose methodologies are available for review. Such methodologies provide the manner of biometric data collection, matching, and reporting, as well as assumptions surrounding the test effort.

Deleted: Biometric t

Deleted: , particularly on a large-scale database

Deleted: s

Deleted: published

A limitation of leveraging biometric test data to evaluate technology suitability for 1:300m identification is the inability to directly extrapolate test performance to real-world performance. Innumerable factors in fielded systems – such as population composition, acquisition and matching methods, system costs, deployer business and functional requirements, usage environment, and deployment/enrollment timeframe – can have a direct and pronounced impact on performance difficult to measure in biometric tests. This limitation is compounded by the size of the application under consideration. No biometric tests have been executed of sufficient size and methodological applicability to draw direct and definitive conclusions on the use of biometrics in a 1:300m system.

Deleted: pronounced

Deleted: simple fact of

Deleted: :

Deleted: n

3.2.2. Identify Relevant Face, Fingerprint, and Iris Deployments

Results from real-world 1:N deployments provide an indicator of the ability of biometric technologies to perform large-scale identification in operational environments. A wide range of factors, such as methods of biometric data collection and environmental factors, impact the performance of biometric technologies as deployed. These factors are difficult to emulate in controlled test environments, such that real-world performance may differ substantially from controlled test performance. By reviewing current, large-scale facial recognition, fingerprint, and iris recognition implementations, we are able to gauge the degree to which technologies are deployed in large-scale identification systems as well as the typical success rates of such deployments.

However, balanced performance data is difficult to gather through real-world deployments, due to the difficulty of isolating and controlling variables in specific usage environments (such as multiple enrollee interactions with a biometric system). Not all biometric performance metrics can be acquired through evaluation of large-scale deployments. Specifically, false non-match rates, by definition, are difficult to assess in an operational system. In many deployments only partial information on system scale and performance is made public, due to the potential risk of security breaches or disclosure of suboptimal performance. Therefore quantifiable information from such deployments may be very limited.

This Report incorporates information on biometric deployments as collected through public announcements, publicly disclosed projects, and other public sources.

3.2.3. Interface With Technology Providers

Based on its knowledge of biometric technology providers whose solutions are used in large-scale 1:N identification, IBG has issued questionnaires to a number of biometric developers, vendors, and integrators in order to gather their feedback on issues central to the execution of 1:300m matching[6].

The developers, vendors, and integrators of the core technologies used in 1:N identification are likely to bring substantial insights into the viability of performing identification at a larger scale than has been attempted to date. While their insights will be limited to a specific core technology or set of technologies, information provided on accuracy and scalability, as well as impediments to large-scale deployment, will complement the information gathering

[6] See Appendix III for the text of this questionnaire.

Deleted: Review
Deleted: Large-Scale
Deleted: ial Recognition
Deleted: A
Deleted: Recognition
Deleted: imposter
Deleted: ;
Deleted: s
Deleted: r
Deleted: the useful
Deleted: derived
Deleted: the provides of
Deleted: for
Deleted: is
Deleted: ing
Formatted: Bullets and Numbering
Deleted: AAMVAUID9
Deleted: Draft 1.0

through evaluation of tests and deployments. A clear limitation on the utility of this data is the risk that vendors may provide excessively optimistic, ungrounded, non-reflective, or unrealistic performance estimates. In most cases no method will be at hand to qualify the usefulness of vendor claims.

3.3. Challenges and Limitations

Data collection and analysis pertaining to 1:300m identification faces numerous challenges and limitations related to the overwhelming scale of the identification task. The lack of biometric databases approaching this size, as well as the lack of biometric tests whose results are readily extensible to a 300m record database, are primary challenges in determining if biometrics can be used successfully to perform 1:N matching in a system of 300m records.

The largest biometric database is the FBI's Integrated AFIS (Automated Fingerprint Identification System). IAFIS houses approximately 55m criminal fingerprint records, most of which contain 20 fingerprint images. Other criminal fingerprint databases have been populated over a period of years to include over 12m records. The largest civil ID (i.e. non-criminal) application incorporates 60m records, each with 6 fingerprint images. The largest facial recognition database is approximately 13m records, with a handful of databases containing between 1m and 10m records.

Another challenge in collecting data for evaluation of systems in 1:300m matching is the fact that no vendor has ever had to address in an actual deployment (or in response to a request for proposal) the challenge of 1:300m identification. Vendors therefore may not have reliable or useful information on the degree to which their technology can scale to this level, aside from extrapolations based on performance on smaller databases. The hypothetical nature of the 1:300m question is such that unanticipated performance considerations may come into play once one reaches very large database sizes. Projecting performance from smaller operational or test databases may thus be inconsistent with real-world performance.

Lastly, technology providers with a business focus on addressing issues involved in real-world systems may struggle to project performance in a system in which the real-world constraints around matching and acquisition are not defined. It is broadly understood that biometric data quality has a major impact on enrollment and matching capabilities. It is further understood that the manner of biometric data acquisition – including the equipment used and processes involved – directly impacts data quality. Lastly, technology providers

Deleted: ¶

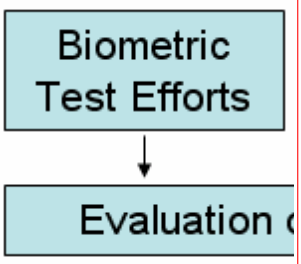


Figure 1: Data Collection Methodology¶

Page Break

Deleted: million

Deleted:

Deleted: ,

Deleted: which

Deleted: 40

Deleted: and 40m civil (background check) fingerprint records

Deleted: 0

Deleted: Facial recognition databases differ from fingerprint and iris databases in that they are capable of searching on legacy facial images not specifically collected for biometric usage. The ability to perform 1:N matching without a dedicated enrollment process allows for the near-immediate generation of facial recognition databases wherever suitable facial images are housed. Regardless, the largest facial recognition databases are measured in the millions, not the hundreds of millions, of records.

Deleted: ,

Deleted: ,

Deleted: purely mathematical

Deleted: theoretical

Deleted: that lacks

Deleted: AAMVAUID9

Deleted: Draft 1.0

always consider performance in conjunction with cost, and must strike balances between optimal performance and the need to fund biometric system development and deployment. As a result, from a technology provider's perspective, contemplating 1:300m identification without also considering data quality, acquisition, and cost factors that so directly impact matching may be seen as an excessively academic exercise for which few reasonable answers can be provided.

Deleted: ¶
As a result,

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

4. Biometric Technologies Capable of 1:N Matching

Facial recognition, fingerprint, and iris recognition are the three technologies evaluated for suitability for a 1:300m application.

Not included in this list of 1:N technologies are speaker verification and DNA verification. Speaker verification technology can be used for small-scale identification, on the order of dozens or hundreds of records, and has been leveraged for criminal forensics-style applications. However we are currently unaware of any deployments in which the technology has been deployed to databases of tens or hundreds of thousands of enrollees. In addition, the behavioral nature of speaker verification is such that attempts to mask or alter one's voice in a fashion not perceptible by third parties may further undermine the ability to execute 1:N identification. As data emerges from studies or deployments in which speaker verification performs large-scale identification, the technology may warrant further evaluation. DNA is not yet developed as an *automated* identification technology in the biometric identification sense; manual processing is still necessary. It can also be argued that DNA identification is based less on measurement of a characteristic than on the material retention of a physical sample. However, DNA identification technology may eventually come to resemble biometric identification as currently implemented, such that the technology could be used to perform large-scale automated identification in DL/ID applications.

The following discussion of facial recognition, fingerprint, and iris recognition technologies' core operations, strengths, weaknesses, and maturity provides a framework for subsequent discussions of tests, deployments, and provider data. This section also discusses multimodal solutions, providing a framework for understanding how biometrics can be deployed multimodally.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

4.1. Facial Recognition

Facial recognition technology is based on features such as the location and composition of distinctive features of the face, as well as the spatial interrelation between the eyes, nose, mouth, cheekbones, chin, and forehead.

Strengths for 1:300m Identification
• High availability reduces FTER (provides near-universal enrollment) • Ability to use multiple matching algorithms to execute 1:N matching; matching algorithms are normally sensor-independent • History of 1:N deployment against large databases • Little to no training required to utilize acquisition devices • Capable of rapid 1:N searches • Compatible with databases of facial images
Weaknesses for 1:300m Identification
• Accuracy issues: distance, angles, lighting, time all impact technology's ability to identify individuals • Changes in hairstyle, facial hair reduce matching capabilities • Performance differs by ethnicity • 1:N searches generally result in candidate lists which must be manually reviewed

Figure 5: Facial Recognition in Identification Systems

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.1.1. Typical Uses

Facial recognition is deployed in large-scale civil ID applications (such as drivers' licensing and voter registration), surveillance applications, law enforcement applications (such as booking stations), and casinos. It is most often deployed in 1:N applications, searching databases of facial images for close matches and returning lists of likely suspects. Increased use of facial recognition has occurred in large-scale ID projects in which facial imaging already takes place and the technology can leverage existing processes and data.

Formatted: Bullets and Numbering

4.1.2. Facial Recognition Processes

4.1.2.1. Acquisition

Facial recognition technology can acquire faces from almost any static camera or video system that acquires sufficient quality and resolution images. For optimal performance, images will be acquired through high-resolution cameras, with users directly facing the camera and with low-intensity frontal lighting of the face, as is consistent with 1:N multiple enrollment detection systems. Such systems utilize controlled and consistent enrollment environments: users are required to stand or sit at a fixed distance from a camera, with fixed lighting and a fixed background. Facial recognition's significant advantage in a DL/ID

Deleted: AAMVAUID9

Deleted: Draft 1.0

environment is its ability to be acquired as part of an existing process (facial photography). The technology's near zero-effort acquisition ensures near total enrollment with almost no impact on current acquisition processes. Facial recognition databases differ from fingerprint and iris databases in that they are capable of creating databases from facial images not specifically collected for biometric usage. The ability to perform 1:N matching without a dedicated enrollment process allows for the near-immediate generation of facial recognition databases wherever suitable facial images are housed.

Formatted: Bullets and Numbering

4.1.2.2. Image Processing

Most facial recognition image processing entails cropping and normalizing images to a consistent size and orientation, converting color images to black and white, and in many cases locating landmarks to assist in automated matching. Characteristics such as the middle of the eyes are used as points of reference. Since most facial recognition systems acquire multiple face images to enroll individuals – from as few as three images to well over 100, depending on the vendor and the matching method – rapid image processing routines are essential to system operations.

Formatted: Bullets and Numbering

4.1.2.3. Distinctive Characteristics

The features most often utilized in facial recognition systems are those least likely to change significantly over time: upper ridges of the eye sockets, areas around the cheekbones, sides of the mouth, nose shape, and the position of major features relative to each other. Areas susceptible to change or obscuration - such as areas of the face immediately adjacent to a hairline - are usually not relied upon for identification. One of the challenges involved in facial recognition technology is that the face is a reasonably variable physiological characteristic. As opposed to a fingerprint, which might be scarred but is difficult to alter dramatically, or the iris, which is reputedly stable for one's entire life, faces can be changed voluntarily enough to reduce a system's matching accuracy. A user who smiles during enrollment and grimaces during verification or identification is more likely to be rejected than one who does not intentionally alter his or her expression during authentication. Behavioral changes such as alteration of hairstyle, changes in makeup, growing or shaving facial hair, adding or removing eyeglasses can impact the ability of facial recognition systems to locate distinctive features.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

4.1.2.4. Template Generation and Matching

Enrollment templates are normally created from multiple processed facial images, although a single image can be used. Multiple facial images are preferred for generation of enrollment templates, as this provides an increased opportunity to "match" the facial positioning of enrollment with that of identification. Templates can vary in size from less than 100 bytes for rapid searching to over three kilobytes. Instead of a single record being returned, in most cases a candidate list with a number of potential matches is returned in large-scale facial recognition identification. For example, a system may be configured to return the ten most likely matches on a search of a 10,000-person database. A human operator would then determine which if any of the ten potential matches were actual matches. The degree to which system performance varies when using a single facial image (as is consistent with DL/ID concepts of operation), as opposed to several facial images acquired with slightly different facial aspects, has not to our knowledge been thoroughly studied.

Facial recognition has been shown susceptible to high error rates, particularly FNMR, when any of several variables are introduced into the matching process. Such variables include the time lapsed between enrollment and matching, the angle of facial acquisition, lighting, distance, facial hair, and glasses.

Formatted: Bullets and Numbering

4.1.3. Competing Facial Recognition Technologies

A handful of facial recognition technologies compete within the biometric market. Solutions may be based on global features, leveraging the common eigenfaces method of rendering facial images by combining features from a database of facial images. Other solutions are based more directly on localized features. Other solutions are based on spatial ratios and relationships between certain fixed points on the face.

Efforts are currently underway within U.S. and international standards groups (ICAO TAG NTWG, ISO/ IET JTC1 SC 37, INCITS M1 Biometrics) to standardize the manner in which facial images are acquired and utilized for biometric matching. When using standard image capture and landmark location processes (e.g. positioning the eyes at a fixed distance for all users), images are more likely to be usable by multiple matching algorithms.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

4.2. Fingerprint

Fingerprint biometrics are based on the ridges, valleys, ridge endings, loops, whorls, and other distinctive features found on the human fingerprint. Fingerprint systems used in 1:N applications, the focus of this Report, are referred to as AFIS (Automated Fingerprint Identification Systems).

Strengths for 1:300m Identification
• Based on distinctive characteristics • Availability of multiple samples (up to ten fingerprints) increases overall accuracy and scalability • Mature and widely accepted standards in DPI (dots per inch), resolution provide framework for consistent technology usage • Strong competition in market drives emergence of new solutions
Weaknesses for 1:300m Identification
• Fingerprint quality varies by age, race; subject to wear and tear (both incidental and intentional) • Individuals with low-quality prints are likely to have several low-quality fingerprints • Fingerprint correlation: individuals' fingerprints (index, middle, ring) share global and feature-based characteristics, and are not completely independent • Percentage of users unable to enroll due to fingerprint quality • Accuracy can diminish over time • Sensor surfaces can be scratched, require maintenance • Scalability of multiple-finger civil AFIS (flat fingerprint) systems uncertain • Highly contingent on quality of initial data

Figure 6: Fingerprint in Identification Systems

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.2.1. Typical Uses

Fingerprint technology is used by hundreds of thousands of people daily to verify availability for public services, to access networks and PCs, to enter restricted areas, and to authorize transactions. The technology is used broadly in a range of vertical markets and within a range of horizontal applications, primarily PC / Network Access, Physical Access / Time and Attendance, Civil ID, and Criminal ID.

Including original equipment manufacturers and application developers, there are over 100 companies operating in the fingerprint marketplace. Approximately half of these companies are "core technology" firms, meaning that they manufacture or develop a basic component of the fingerprint system such as a sensor or an algorithm. These companies may utilize a proprietary sensor technology or use sensors from another manufacturer. In addition, these core technology companies may develop or manufacture the following components:

Deleted: AAMVAUID9

Deleted: Draft 1.0

- Sensors and modules (designed for integration into third party devices or peripherals)
- Devices and peripherals (ready-to-deploy units for logical and physical access)
- Algorithms (perform extraction and matching functions)
- Application software (enable PC or network access)

Fingerprint technology can also be divided according to minutia-based and pattern-matching vendors. Approximately 75% of fingerprint solutions use minutia-based extraction algorithms, which generate and compare templates based on the relative position and direction of dozens of ridge endings and bifurcations found in fingerprints. Pattern matching algorithms are based on the regional characteristic present across multiple ridges as opposed to single points.

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.2.2. AFIS Technology

AFIS technology utilizes groups of fingerprints, as well as fingerprint classification methods, to enable large-scale searches against databases of thousands to millions of fingerprints. AFIS was first developed in the mid to late 1970's as a means of automating manual searches required to identify a fingerprint from the tens of millions of fingerprints in FBI files. Today, the term "AFIS" refers to the *technology* used to automate large-scale fingerprint searches as well as the *industry* inclusive of automated acquisition, storage, and identification of fingerprints.

From a technology perspective, the AFIS industry can be viewed as two distinct segments: (1) fingerprint acquisition and (2) fingerprint storage and processing. Fingerprint acquisition components collect high-quality electronic fingerprint images, either directly from placement of fingerprints on "live-scan" readers or from scanned ink cards. Fingerprint storage and processing components generate fingerprint templates and perform 1:N matching against fingerprint databases. Mature standards define the capture resolution, permissible distortion, and compression ratios of fingerprint images, ensuring interoperability of images across AFIS vendor systems.

Formatted: Bullets and Numbering

4.2.3. Flat and Rolled Fingerprints in AFIS Deployments

AFIS deployments can utilize either rolled or flat fingerprints. While the same underlying matching technologies can be used in both rolled-print and flat-print matching, the two types of solutions differ in terms of performance and usability.

Deleted: AAMVAUID9

Deleted: Draft 1.0

Rolled fingerprints are images of the entire fingerprint space from "nail to nail". To acquire rolled fingerprint images, an operator must manually roll the finger across a flat surface throughout the approximately 180 degrees of the fingerprint region. This process acquires a larger quantity of fingerprint data usable for 1:N matching than does flat fingerprinting. However, acquiring rolled fingerprints is a time-consuming, labor-intensive process. Depending on the skill of the operator, the dexterity of the enrollee, and the equipment used, each fingerprint may require several seconds to image. In addition, based on the state of current fingerprinting technology, an operator must physically manipulate the finger in order to acquire a rolled fingerprint – the process is too difficult, and the parameters for acceptable images too narrow, to allow enrollees to provide data simply based on instructions or verbal guidance. Rolled fingerprint images are used in criminal booking processes, employment background check applications, and any situation in which comparison against law enforcement databases is a requirement.

Flat fingerprints are images of the central area of the fingerprint, acquired through placement of the fingerprint onto a flat surface. Acquiring flat fingerprints is a faster and simpler process than acquiring rolled prints, although flat fingerprint images provide a smaller quantity of biometric data than rolled images. In addition to use in a wide range of 1:1 applications, flat fingerprints have historically been used in 1:N civil ID applications such as national ID programs, card issuance programs, and social security and benefits programs.

While rolled fingerprint systems are inherently more scalable and accurate than flat fingerprint systems due to the fact that they are based on a larger quantity of distinctive data per individual, there are as yet no independent studies that demonstrate precisely how much more accurate rolled fingerprints are than flat. This is a critical question for large-scale applications, as the improvement in performance for a given number of fingerprints must be sufficient to warrant the additional effort required for rolled fingerprint acquisition.

Formatted: Bullets and Numbering

4.2.4. Fingerprint Processes

4.2.4.1. Acquisition

Image acquisition is a major challenge for fingerprint providers, as fingerprint quality varies substantially from person to person and from fingerprint to fingerprint. Certain populations are more likely than others to possess faint or difficult-to-acquire fingerprints, whether due to occupational wear and tear or physiology. In addition, environmental factors can impact

Deleted: AAMVAUID9

Deleted: Draft 1.0

image acquisition: in cold weather, fingerprint oils (which improve image quality) dry up, such that fingerprint images may appear faint. Sensor size can also impact a system's accuracy and performance. Very small sensors acquire a smaller portion of the fingerprint, such that less data is available to enroll and match templates. Users with large fingers may find it difficult to place their fingerprint in a consistent fashion, leading to false non-matches.

Formatted: Bullets and Numbering

4.2.4.2. Image Processing

Image processing subroutines convert the fingerprint image's gray pixels to white and black. A series of thick black ridges (the raised part of the fingerprint) results, contrasted to white valleys. The ridges are then "thinned" down to a single pixel in width to enable precise location of features.

Formatted: Bullets and Numbering

4.2.4.3. Location of Distinctive Characteristics

The fingerprint is comprised of ridges and valleys which form distinctive patterns, such as swirls, loops, and arches. Many fingerprints also have a core, a central point around which these patterns curve. Points found at the lower left or right corner of the fingerprint around which ridges are centered in a triangular shape are known as deltas. Discontinuities and irregularities in ridges and valleys – known as minutiae – are the features upon which most fingerprint technologies are based. The primary minutia types are ridge endings (the point at which a ridge ends) and bifurcations (the point at which a ridge splits). Depending on the size of the sensor and the sensitivity of the algorithm, a typical fingerprint may produce between 30 and 50 minutia – larger platens acquire more of the fingerprint image such that more minutiae can be located. This information is fairly stable throughout one's life, and differs from fingerprint to fingerprint.

Fingerprints contain sufficient information to enable large-scale identification using multiple fingerprints. However, conducting such large-scale searches is time consuming and processor-intensive. To limit the number of fingerprints that must be searched in identification systems (and thereby to limit search time and processing demands), AFIS technology also classifies the group of fingerprints acquired from each individual according to prints' global characteristics. An AFIS may therefore only need to search that percentage of records whose group classification matches that of the enrollee, instead of searching an entire fingerprint database. The percentage of an AFIS database that must be searched subsequent to classification is referred to as a "penetration rate" – lower penetration rates result in faster searches. Collecting more fingerprints per individual results in a lower

Deleted: AAMVAUID9

Deleted: Draft 1.0

penetration rate. However, it is critical that fingerprints be classified correctly, or else a multiple enrollee may go undetected in a 1:N search. Newer AFIS systems utilize matching methods that do not rely on this type of classification scheme.

4.2.4.4. Template Generation and Matching

Vendors utilize proprietary algorithms to locate fingerprint minutiae. Information used when mapping minutiae can include the location and angle of a minutia point, the type and quality of minutia, and the distance and position of minutiae relative to the core. Fingerprint images may contain distortions and "false minutiae" that must be filtered out before template creation; scars, sweat, and dirt can appear as minutiae. Algorithms scan images and eliminate anomalous features that seem to be in the wrong place, such as adjacent minutiae or a ridge crossing perpendicular to a series of other ridges. A large percentage of false minutiae are discarded in this process, ensuring that the template generated for enrollment or matching accurately reflects legitimate biometric data.

Fingerprint templates can range in size from approximately 200 bytes to over 1000 bytes (in contrast to the 10-20 kilobytes required to store a single compressed fingerprint image). These templates cannot be "read" like a fingerprint image. Matching algorithms are required to process templates and to determine the correlation between the two.

4.2.5. Sensor types

Fingerprints are acquired through optical, silicon, and ultrasonic sensors.

Optical technology is the oldest fingerprint imaging technology and the most widely used for 1:N systems. Optical technology has several strengths: proven reliability over time, resistance to electrostatic discharge, resolutions of 500 dpi and above, and large surface area (certain optical sensors can acquire multiple fingerprints in a single placement). Weaknesses include size and power constraints – the platen requires more surface area and depth than silicon technology, such that optical technology cannot e built into very small devices – and difficulty acquiring dry fingerprint images. Optical technology is widely deployed in high-traffic public sector applications such as DL/ID.

Silicon technology uses coated chips to image fingerprints. Most silicon fingerprint technology is based on capacitance, wherein the silicon sensor acts as one plate of a capacitor and the finger acts as the second plate. Silicon sensor strengths include image quality (approaching that of many optical devices), modest size and power requirements

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

(such that sensors can be integrated into small, low-power devices), and low cost. Silicon sensor weaknesses include durability, susceptibility to electrostatic damage, and performance in challenging conditions. Silicon sensors are primarily deployed as logical access solutions, have found limited but increasing deployment in physical access deployments, and are rarely deployed in high-traffic public sector applications such as DL/ID.

Ultrasonic devices generate fingerprint images by emitting ultrasonic waves, measuring the "echo" returned when the acoustic waves meet the ridges and valleys of the fingerprint. Among the advantages of this method are that ultrasonic devices are better able to acquire images from low-quality fingerprints with surface contaminants – dirt, grease – that normally reduce image quality. Theoretically, better image quality will result in more accurate fingerprint matching, though there have been no tests that demonstrate ultrasonic scanner's compatibility with multiple vendor's matching algorithms. The platen used in ultrasonic technology does not require a special coating, as is often the case in optical imaging. Disadvantages of the ultrasonic imaging method include a bulky device size, the presence of moving parts necessary to image the fingerprint, as well as the length of time required to gather images during enrollment. Ultrasonic technology can be used in civil ID applications such as DL/ID.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

4.3. Iris Recognition

Iris recognition technology is based on the ridges, furrows, and striations that characterize irises.

Strengths for 1:300m Identification
• Extremely low False Match Rates • Thought to be a highly stable and distinctive physiological characteristic, unchanging over time • Ability to leverage both irises, which have been shown in vendor tests to be highly independent
Weaknesses for 1:300m Identification
• Scalability in 1:N applications not yet proven in field evaluations • Enrollment can be problematic for certain users • Acquisition of iris can be time consuming and difficult

Figure 7: Iris Recognition in Identification Systems

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.3.1. Overview

The iris is a highly distinctive characteristic, reportedly stable at birth and unchanging through one's life, and has not been shown to be alterable in any fashion (although the appearance of an iris can possibly be masked through use of certain contact lenses). Iris recognition technology is primarily deployed in high-security physical access implementations, but has found increasing acceptance in travel and transportation and civil ID applications. One company (Iridian) holds the key patents for utilizing the iris for identification. Iridian both licenses its core technology to integrators and developers and it markets systems directly to deployers. Although alternative implementations of iris recognition technology have recently begun to emerge, Iridian has historically been very aggressive about defending its patent rights.

Formatted: Bullets and Numbering

4.3.2. Iris Recognition Processes

4.3.2.1. Image Acquisition

Iris recognition matching requires the acquisition of a high-resolution image of the eye, illuminated by an infrared imager, in order to effectively map the details of the iris. The acquisition process, and the effort required on the part of the user, differs according to the type of acquisition device used. Primary iris recognition imaging systems include kiosk-based systems, physical access devices using motorized cameras, and inexpensive desktop cameras. Although iris recognition vendors do not emphasize their use of infrared

Deleted: AAMVAUID9

Deleted: Draft 1.0

light, each system does rely on infrared imaging using wavelengths in the 700-900nm range (judged to be safe by the American Academy of Ophthalmology).

Depending on the quality and positioning of the acquisition device, and the level of training and supervision granted to the enrollee, acquisition of iris images requires moderate to high levels of training and attentiveness. Users must be cognizant of the manner in which they interact with the system, as enrollment and matching require fairly precise positioning of the head and eyes. Also, users with poor eyesight, or those incapable of lining up their eye with the technology's guidance components, have difficulty using the technology.

Formatted: Bullets and Numbering

4.3.2.2. Image Processing

After the eye is located, algorithms locate the iris' outer and inner borders. Locating the iris-pupil border can be challenging for users with very dark eyes, as there may be very little difference in shade as rendered through 8-bit grayscale imaging. Once the parameters of the iris have been defined, a black and white image of the iris is used for feature extraction. The core technology can account for pupil dilation, eyelid occlusion, and reflections due to the acquisition camera. When the pupil dilates, the iris patterns shrink and expand in a normalized fashion such that algorithms can translate a dilated match to a non-dilated enrollment.

Formatted: Bullets and Numbering

4.3.2.3. Distinctive Features

Iris recognition technology uses a horizontal band extending from the far left to the far right of the iris for feature extraction. The patterns that comprise the visual component of the iris are highly distinctive. The trabecular meshwork, a tissue that gives the appearance of dividing the iris in a radial fashion, is the primary distinguishing characteristic. Other visible characteristics include rings, furrows, freckles, and the corona. Tests have shown that individuals' left and right eyes have different iris patterns, and that even identical twins' irises have almost no statistical similarity. Iris recognition algorithms map segments of the iris into hundreds of independent vectors. The characteristics derived from iris features are the orientation and spatial frequency of distinctive areas along with the position of these areas.

Formatted: Bullets and Numbering

4.3.2.4. Template Generation and Matching

The vectors located by the iris recognition algorithm are used to form enrollment and match templates, which are generated in hexadecimal format as opposed to binary. Depending on the iris recognition solution, between one and four iris images may need to be captured for enrollment template generation. The use of multiple images ensures that the data extracted

Deleted: AAMVAUID9

Deleted: Draft 1.0

to form a template is consistent. Iris recognition solutions generally perform identification as opposed to verification, meaning that the match template is compared against multiple enrollments until a match is located. The matching process is very rapid, with hundreds of thousands of records searched per second.

Iris recognition is unique among the three biometrics under consideration in that it is not designed to generate candidate lists. 1:N searches stop once a template is encountered that meets the required threshold, with the implicit assumption that no other template in the database would have been a better match. This approach has grown out of the technology's history of exceptionally low FMR – continuing a search after a match is found would be seen as unnecessary.

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

4.4. Multimodal Solutions

The challenges involved in scaling biometric systems to execute 1:300m identification in a DL/ID environment may be mitigated through the use of multimodal biometric systems. By leveraging combinations of technologies such as facial recognition, fingerprint, and iris recognition, multimodal systems may be capable of providing a higher degree of accurate scalability than any single biometric technology. Such an approach may be the only solution to the 1:300m challenge. In order to determine if and how multimodal systems can be used to improve accuracy and enrollment capabilities for the purposes of 1:300m identification, it is necessary to understand the basic operations and categorizations of multimodal systems.

Formatted: Bullets and Numbering

4.4.1. Definition

Multimodal biometric systems are those which utilize, or collect for the purpose of utilizing, more than one physiological or behavioral characteristic for enrollment, verification, and/or identification. Multimodal systems address specific problems found in monomodal biometric systems, including the following:

Formatted: Bullets and Numbering

- Biometric systems are subject to false match and false non-match errors. Depending on the type of biometric system deployed, excessive matching errors can lead to security breaches, undetected fraud, and processing delays.
- Biometric systems are subject to failure to enroll and failure to acquire errors. Such errors can be attributable to lack of required physiological characteristics, to insufficiently distinctive biometric characteristics, or to an inability to adhere to device interaction requirements. FTE and FTA errors result in a percentage of individuals permanently or temporarily unable to use a given biometric system. This creates problems for deployers, as a backup authentication method must be maintained, and malicious users may intentionally fail to enroll in order to attack the weaker authentication method (e.g. password).
- Recent demonstrations have shown that many biometric systems can be fooled by fake fingerprints, some with little effort, others with substantial effort. This raises the possibility that difficult-to-repudiate transactions could be created and associated with an individual without his awareness, and that individuals could easily circumvent 1:N detection through use of fraudulent data.

Deleted: AAMVAUID9

Deleted: Draft 1.0

Multimodal biometric systems are designed to provide the following benefits:

1. **Reducing false non-match rates and false match rates.** By deploying more than one biometric technology for 1:N and/or 1:1 processing, and intelligently combining or fusing match results from both systems, it may be possible to reduce the overall system's matching error rates. For example, to reduce 1:1 false matching, a multimodal system can provide two subsystems against which a user must match in order to defeat the system. Similarly, to reduce 1:N false non-matching, a multimodal system may only require that a user match against one of two subsystems.
2. **Providing a secondary means of enrollment, verification, and identification for users unable to enroll and/or authenticate through a primary biometric technology.** By deploying more than one biometric technology, a deployer can ensure that a higher percentage of individuals are enrolled and matched in a biometric system, reducing the need for fallback or secondary processing. This in turn can reduce security risks.
3. **Combating attempts to spoof biometric systems through non-live data sources such as fake fingers.** By implementing a multimodal system that requires dual authentication, an attacker must successfully spoof both systems, or spoof one while attempting to match as a multiple enrollee in another.

These direct benefits can lead to indirect benefits such as increased system security, reduced deployment costs, and increased ease of use. However, realizing benefits from a multimodal biometric system requires intelligent combination and utilization of technologies, devices, and algorithms; requires an application for which a multimodal solution is both viable and useful; and requires effective design of enrollment and matching processes. Issues and variables involved in designing multimodal biometric processes, selecting multimodal technologies, and identifying suitable deployment environments for multimodal solutions are addressed in the following sections.

4.4.2. Concepts of Operations

Multimodal biometric systems can take a variety of forms with wide-ranging levels of complexity. In order to provide a method of categorizing the major functional elements of multimodal systems, and to allow for a common baseline for discussions of multimodal concepts of operation, a five-stage framework, incorporating (acquisition, matching, output,

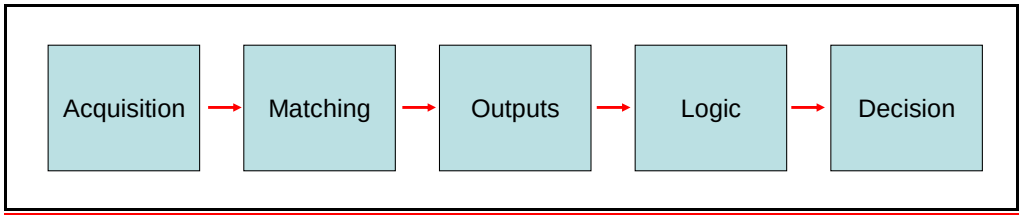
Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

logic, and decision, can be used. Variables within multimodal biometric systems can be categorized according to the major steps in the biometric processing cycle:



- **Acquisition Process** variables apply to collection and assessment of biometric data in a 1:300m system.
- **Matching Process** variables apply to comparison of biometric data in a 1:300m system.
- **Output Process** variables apply to generation of match results through biometric comparisons in a 1:300m system.
- **Logic Process** variables apply to methods of utilizing one or more sets of match results in a 1:300m system.
- **Decision Process** variables apply to decisions that result from biometric transactions or search events in a 1:300m system.

This approach allows us to (1) isolate areas that inform multimodal performance and (2) allow for inclusion of various modes of multimodal system implementation, from simple to complex.

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.4.2.1. Multimodal Acquisition

A wide range of variables apply to the collection of biometric data during enrollment and/or identification. This collection event, and the parameters around the event, falls under the auspices of *Acquisition* variables. In nearly all multimodal biometric systems more than one acquisition device will be necessary, as each biometric discipline utilizes its own acquisition technology: cameras, scanners, microphones, etc. Although a multimodal system may be designed such that a primary biometric is used for most transactions and a secondary biometric is used for exception cases, in a DL/ID environment, it is assumed that both systems will be utilized. This directly impacts the manner of acquisition of biometric data. *Acquisition* variables are particularly important due to their direct impact on enrollees and end users. These variables may impact transaction times, storage requirements for both capacity and flexibility, and level of supervision required for system operation. *Acquisition*

Deleted: AAMVAUID9

Deleted: Draft 1.0

variables are also important due to their relation to limiting FTER. With sufficient time and effort, an enrollment operator can drive FTER down to near-zero for many technologies; however, this does not address the needs of a transactional system in which a user may need to provide data in a time-constrained, unattended usage environment. In this environment, users may enroll successfully but be incapable of authenticating on subsequent interactions with the system. The following table describes the primary *Acquisition* variables and lists ranges associated with these variables.

Acquisition Variable	Description	Range
1. Number of Acquisition Devices	**Number of acquisition devices involved [maximum].** The number of acquisition devices in a multimodal system can impact system costs and deployment footprint.	1-N devices utilized
2. Number of Core Technologies	**Number of core technologies involved [maximum].** In most cases, each core technology (e.g. fingerprint, facial recognition) will be associated with a single acquisition device. However in some cases a single device may be capable of acquiring more than one biometric type, e.g. iris and face.	1-N technologies utilized

Figure 8: Acquisition Variables

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.4.2.2. Multimodal Matching

Nearly all multimodal systems incorporate multimodal *Matching Processes*. The primary exception is a multimodal system in which more than one biometric is acquired on enrollment but only one biometric is used on an ongoing basis for 1:N and 1:1 functionality. Matching is separate from output and logic, being applicable solely to matching processes. In almost all cases the protocols and mechanics associated with template generation and comparison are not known beyond the vendor. The following table describes the primary *Matching* variables and lists ranges associated with these variables.

Matching Variable	Description	Range
1. Number Of Core Technologies	**The number of core technologies involved.** Multimodal matching processes are directly impacted by the number of core technologies involved. Because biometric matching is, for many technologies, a processor-intensive function, multimodal systems may be designed to only match secondary biometrics in certain situations. A system in which three different biometrics are available may only match two.	1-N technologies utilized
2. Number Of Algorithms	**The number of matching algorithms involved.** Multimodal systems must, with rare exceptions, utilize a different matching algorithm for each biometric characteristic acquired (e.g. face, finger). At the same time, multimodal systems can be designed to utilize a single matching algorithm for each biometric characteristic, or may utilize multiple algorithms to process a given piece of biometric data. This multi-algorithmic approach in itself contains many potential sub-	1-N matching algorithms utilized

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

	variables, including weighting of match scores at the logic processes phase. Most systems do not apply a multi-algorithmic approach to a single technology.	
3. Source Fusion	**A process in which biometric samples (e.g. identifiable images, voice patterns) are combined in either raw or processed form to create a new meta-sample; this meta-sample may be processed through a different algorithm than either source sample.** Source fusion is a novel process by which more than one biometric type is fused at the sample stage in order to generate a "new" biometric. In most cases a pattern-matching algorithm would be used to process this new type of biometric data.	Data fused from characteristics 1, 2,…N

Figure 9: Matching Variables

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.4.2.3. Multimodal Output

Outputs that result from matches in multimodal systems are identical to those in monomodal systems. For 1:300m identification systems outputs include Rank 1 and Rank N results. Match responses from each component of a multimodal system are critical to developing large-scale ID systems such as in a DL/ID environment. The following table describes the primary *Output* variables, and lists ranges associated with these variables.

Output Variable	Description	Range
1. Rank 1	**Match output rendered as a single candidate.** Open and closed set multimodal systems are each capable of generating *Rank 1* Match outputs, returning a Candidate ID corresponding to the record with the strongest correlation score.	Candidate ID N
2. Rank N	**Match output rendered as a list of candidates.** Open and closed set multimodal systems are each capable of generating candidate lists indicating the top N matches, or *Rank N matches*, each of which may be associated with a given score or simply ranked in numerical order.	Candidate ID 1, 2, 3…N

Figure 10: Output Variables

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.4.2.4. Multimodal Logic

Multimodal *Logic* is the focus of considerable active multimodal research, as academics and vendors investigate the most effective method of combining results from multiple systems to improve overall biometric system performance. Determining the most effective way of combining match results from disparate systems is at the core of multimodal systems operation. To date, fielded products have often used rudimentary *and/or* logic in multimodal systems, driven partly by a desire to simplify deployment but also by a lack of strong familiarity with what components' match scores truly indicate. More advanced approaches have been developed by vendors with access to the core intellectual property behind more than one biometric modality and by integrators with experience in developing solutions that

Deleted: AAMVAUID9

Deleted: Draft 1.0

fuse inputs through algorithm layers to generate improved cross-system results.

Several of the following *Logic* areas can be categorized as fusion processes. Fusion systems use specially developed decision models to intelligently combine the results from more than one biometric system into a "master" decision. Fusion biometrics leverage the probabilities associated with biometric match events in driving match/no-match decisions. For example, a strong match on an iris recognition system may require only a very weak match on a parallel-operating facial recognition system, while a weak match on iris recognition may require a much stronger facial recognition match. In order for fusion biometrics to provide demonstrable improvements in accuracy, a developer requires access to detailed information regarding vendor-specific biometric scores: specifically, how scores correspond to probabilities when determining matches and non-matches. The following table describes the primary *Logic Process* variables and lists ranges associated with these variables.

Logic Variable	Description	Range
1. System Weight	**Multimodal logic in which one system's output is weighted more heavily than another system's.** *System Weight Logic* is generally applicable in systems that combine a strong biometric with a weak biometric. In order to ensure that the stronger biometric technology's decision impacts match decisions, a 75/25 or 80/20 weight may be accorded to a given system, such that an extremely low match score in the weaker system would be necessary to cause a non-match. This logic is designed to reduce FNMR and FMR.	Weight A (System 1), Weight B (System 2), …Weight N (System N)
2. Score Weight	**Multimodal logic in which relatively strong match scores figure more heavily in a decision process than low match scores.** *Score Weight Logic* is applicable to multimodal systems that utilize core technologies of roughly equivalent strength, such as fingerprint and iris recognition or facial recognition and voice verification. This logic can only be applied to multimodal systems whose outputs are non-binary (which would include native and standardized match scores as well as probability outputs). This logic is designed to reduce FNMR, built on the assumption that a single strong match should bear a stronger weight in the overall logic.	Weight A (System Score 1), Weight B (System Score 2), … Weight N
3. Combined Score 'And'	**Logic process in which the application requires a total score across all systems.** This logic process allows for a variety of combinations of match scores, often in conjunction with system weight and/or score weight logics, as above. This logic process allows for more detailed means of incorporating results from two or more systems.	Combined Minimum Score N on Systems A, B

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

Logic Variable	Description	Range
4. Match Score 'And'	**Logic process in which the application require a minimum score on all systems.** A more advanced version of Binary Match 'And', this logic process entails that specific match scores known to be associated with desired probability levels be surpassed for each component of a multimodal system. Knowledge of vendor thresholds and how they map to user requirements is a prerequisite of this logic process, designed to reduce FMR.	System A Minimum Score N *and* System B Minimum Score N
5. Rank N Match 'And'	**Logic process in which the system requires a Rank N Candidate on all systems.** *Rank N Match 'And' Logic* establishes a minimum rank for 1:N searches across all systems within which results for a given individual are assumed to be valid. For example, a Rank 5 Match 'Or' system would return all matches in which an individual was within the top 5 for all systems.	Rank N on Systems 1 and 2 and …N
6. Rank 1 Match 'And'	**Logic process in which the system requires a Rank 1 Candidate on all systems.** *Rank 1 Rank 1 Match 'And' Logic* requires that a Rank 1 match be attained on all systems for an individual to be declared a match; this process is designed to reduce system FMR.	Rank 1 on Systems 1 and 2 and …N

Figure 11: Logic Variables

4.4.2.5. Multimodal Decisions

Multimodal *Decisions*, which represent the institution's policies for matching, non-matching, identification, and other functions, are similar to monomodal decision processes with the addition of more complex retry sequences at certain points. Contingent multimodal systems can be designed with 'gates' that implement decision processes subsequent to matching within one biometric technology, with results potentially triggering match decisions or triggering activation of a secondary biometric technology. *Decisions* are closely tied to *Logic*, and represent the institution's actions based on a given match event's ability to successfully address an application's logic requirements. The following table describes the primary *Decision* variables and lists ranges associated with these variables.

Decision Process	Description	Range
1. No Match, Terminate Sequence	**Decision process in which no match has occurred, the biometric sequence terminates.** This decision process logically follows a match sequence in which none of the preconditions for matching established at the logic phase are met. If the user has failed to match for N number of transactions, the entire matching sequence may be terminated and revert to manual authentication or user lockout.	No Match
2. No Match, Retry	**Decision process in which no match has occurred, the biometric sequence is reinitiated.** This decision process is utilized when an individual is granted multiple sequences to authenticate and fails to authenticate within one of the initial sequences.	No Match: Retry N
3. Match, Grant Access	**Decision process in which a match has occurred.** This decision process is a common result in which a user	Match

Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Deleted: AAMVAUID9
Deleted: Draft 1.0

	successfully meets conditions established at the logic phase.	
4. Rank 1 Match Found	**Decision process in which a candidate has been identified as Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which a Rank 1 match has been found to meet the requirements of the 1:N logic. In most cases the matching record would be returned for further action (investigation, grant of access).	Match, ID#
5. Rank 1 Match Not Found	**Decision process in which no candidate has been identified as Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which no Rank 1 match has been found that meets the requirements of the 1:N logic.	No Match
6. Candidate Found Within Ranks 1-N	**Decision process in which a candidate has been identified as being within N positions of Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which an individual is located within N places of Rank 1 in a fashion that meets 1:N logic requirements. In most cases the matching record would be returned for investigation.	Match (rank), ID#
7. Candidate Not Found Within Ranks 1-N	**Decision process in which no candidate is identified as being within N positions of Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which no individual is located within N places of Rank 1 in accordance with 1:N logic requirements.	No Match
8. Inconclusive	**Decision in which the match result is inconclusive to execute a decision.** This decision process may result in a declaration of no match or in a triggering or a retry cycle.	No Match / Retry N

Figure 12: Decision Process Variables

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

4.4.3. Multimodal Technology Combinations

4.4.3.1. Multimodal 1:N Technology Combinations

Potential multimodal technology combinations for 1:300m identification applications include *Fingerprint and Facial Recognition*, *Iris Recognition and Facial Recognition* and then *Fingerprint and Iris Recognition*. Limitations and advantages of these 1:N multimodal technology combinations are as follows.

Formatted: Bullets and Numbering

4.4.3.2. Fingerprint and Facial Recognition

The use of fingerprint and facial recognition as a 1:N multimodal solution provides advantages in terms of ability to leverage legacy databases and (in the case of facial recognition) to leverage existing processes such as photo capture. Because facial recognition has been sanctioned by ICAO as the primary interoperable biometric technology for passport usage, it is likely that substantial effort will be dedicated to developing multimodal solutions that utilize this mandatory data piece in addition to more reliable identifiers such as fingerprint and iris recognition. In terms of 1:N functionality, facial

Deleted: AAMVAUID9

Deleted: Draft 1.0

recognition's ability to function as a gross classifier allows it to reduce the size of large 1:N databases and to effect more rapid 1:N fingerprint searches. Executing parallel full-scale 1:N searches through both fingerprint and facial recognition is unlikely to prove highly beneficial, as the results from the facial recognition will not be reliable, even on small databases.

This underscores a challenge of this technology combination: if fingerprints cannot be obtained from a given subject, then only facial recognition can be utilized, meaning that a much less robust 1:N search would need to be executed. Furthermore, a deployer cannot simply assume that a search is required only on a database of individuals unable to enroll in the fingerprint system. A motivated individual can mar his or her fingerprints such that he or she could fail to enroll after having enrolled previously. Therefore facial recognition must be optimized to work as a standalone system to avoid providing multiple enrollees with simple workarounds.

Formatted: Bullets and Numbering

4.4.3.3. Iris Recognition and Facial Recognition

Iris recognition and facial recognition bring the substantial advantage of being capable of being imaged through a single user process and through a single-housing acquisition device (which may contain two separate imaging elements). Therefore many of the process-driven impediments that face multimodal systems are not present in this technology combination. Narita Airport in Tokyo is testing this exact technology combination to determine the efficacy of the combined technologies. The limitation of this technology combination has to do with the relatively marginal role that facial recognition can play in a system with a strong and highly available biometric such as iris recognition. In most cases, if the iris can be reliably imaged, then the facial recognition components will add very little, such that the cost/benefit of collecting and maintaining such data can be called into question. Also, since iris recognition is always deployed as a day-forward solution as opposed to a solution that leveraged legacy data, there is less need to address existing large-scale databases. In extremely large-scale 1:N systems, facial recognition could serve as a gross classifier to reduce the demands on the 1:N iris search, as iris technology has not been deployed in highly scaled application environments (those with 1m+ enrollees).

Formatted: Bullets and Numbering

4.4.3.4. Fingerprint and Iris Recognition

For systems in which certainty regarding match results is an absolute necessity, fingerprint and iris recognition offer similar capabilities in terms of reliable 1:N matching. Fingerprint is more proven in real-world applications, and has been shown to scale with large loads of

Deleted: AAMVAUID9

Deleted: Draft 1.0

applicants; iris recognition is harder to circumvent than fingerprint technology, is more universally available, and provides high levels of accuracy. Each technology offers multiple samples, increasing scalability and accuracy. It is likely that for many deployers, iris recognition and fingerprint represent similar-enough capabilities that using both will be excessive. Neither is designed to provide the rapid, inexpensive database-reducing 1:N gross search functions of facial recognition.

Deleted: AAMVAUID9

Deleted: Draft 1.0

5. Performance Evaluation Parameters

5.1. Performance Metrics and Error Types

Standardized approaches to evaluating biometric systems are still emerging. Fundamental issues such as how to measure matching performance are the subject of considerable debate. While false match rates and false non-match rates (along with their corollary correct match and correct non-match rates) are the most widely understood and reported performance metrics from biometric systems, a different set of performance metrics may be more capable of capturing all potential outcomes from 1:N searches.

The following discussion presents two approaches available to UID9 in evaluating 1:N performance: *False Match and False Non-Match Rates* as well as *Open Set Identification Performance Metrics. False Match and False Non-Match Rates* are more commonly used to discuss system performance. However, based in its specific needs, UID9 has concluded that *Open Set Identification Performance Metrics* provide a much clearer understanding of real-world performance.

Subsequent to this discussion, failure to enroll rate is addressed. Evaluation parameters such as proven deployment history, response time, and resistance to orchestrated attack are addressed in subsequent Phases.

5.2. False Match and False Non-Match Rates

False match rates and false non match rates enable calculation of the following:

- The likelihood that a first-time enrollee will incorrectly match one or more individuals previously enrolled in a database; also, the likelihood that a first-time enrollee will not be matched against any individuals previously enrolled in a database
- The likelihood that a duplicate enrollee will incorrectly fail to be matched against one or more of his or her records previously enrolled in a database; also, the likelihood that a duplicate enrollee will matched against any or his or her previous enrollments.

Formatted: Bullets and Numbering

Deleted: The primary Phase 1 Report evaluation parameters under consideration within a 1:300m biometric system are false match rates, false non-match rates, and failure to enroll rates

Deleted: False Match Rates

Formatted: Bullets and Numbering

Deleted: 7

Deleted: . These error rates enable calculation of the following:¶
<#>The likelihood that a first-time enrollee will incorrectly match one or more individuals previously enrolled in a database; also, the likelihood that a first-time enrollee will not be matched against any individuals previously enrolled in a database¶
<#>The likelihood that a duplicate enrollee will incorrectly fail to be matched against one or more of his or her records previously enrolled in a database; also, the likelihood that a duplicate enrollee will matched against any or his or her previous enrollments. ¶
<#>The likelihood that an individual will be capable of enrolling in the biometric system; also, the likelihood that an individual will be incapable of enrolling in the biometric system¶
<#>Evaluation parameters such as proven deployment history, response time, and resistance to orchestrated attack are addressed in subsequent Phases. ¶

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

5.2.1. False Match Rates[8]

5.2.1.1. Definition

A biometric system's false match rate (FMR) can be defined in a variety of ways depending on the type of system being evaluated. A universal definition of FMR is

> *The rate of incorrect positive matches by a biometric system for single template comparison attempts.*

In other words, the false match rate represents the projected frequency with which a 1:1 comparison of two biometric records from different individuals will result in a match. This definition is oriented toward biometric verification, in which a user's biometric data is compared against his or her enrolled data. However, in biometric identification systems, each 1:N search may entail comparisons against millions of enrolled users. An exceptionally low 1:1 false match rate would then be required in order to provide effective operations in a 1:N system with 300m users. Therefore the 1:1 false match rate must be evaluated in the context of the number of records to be searched. To illustrate, in a DL/ID system within which 300m users are to be enrolled within a 12-year timeframe, one will eventually approach typical transaction loads of 130,000 enrollees being compared against a 300m person database. This is the equivalent of 39 trillion comparisons in one day.

For purposes of the 1:300m evaluation, the *effective* false match rate is a more useful measure of system effectiveness. The effective false match rate is the projected frequency with which a given customer will be falsely matched against a database in a 1:N search. In order to calculate the effective false match rate, one must know a given technology's 1:1 false match rate, the number of records in the database, and the transaction rate.

Note that biometric technologies may return several possible matches subsequent to a 1:N search; this transaction represents a single false match (assuming that none of the records returned is determined to be a true match), regardless of the number of records incorrectly returned. It may be worth recording the number of records falsely returned in the 1:N search, but effective FMR is a measurement of the percentage of transactions in which match errors occur vs. the percentage in which matching errors do not occur.

[8] Due to the 1:N nature of the technology assessment, error metrics such as "false accept rate" and "false reject rate" are not utilized in this Report, unless when quoting vendors: acceptance and rejection are applicable in 1:1 applications, in which a biometric match normally results in "acceptance" to a protected resource.

Deleted: <#>False match rates are defined in order to establish a frame of reference against which reported false match rates from biometric tests and deployments can be compared. ¶

Deleted: .

Deleted:

Deleted: probability that

Deleted: n

Deleted: ¶

Deleted: 5

Deleted: Note that this match load does not take peak transaction days into account; 130,000 enrollees per day would represent a fairly enrollment rate.

Deleted: probability

Deleted: that

Deleted: would need

Deleted: to

Deleted: and

Deleted: given

Deleted: the size of

Deleted: and

Deleted: ¶

Deleted: for

Deleted: ing

Deleted: a "gallery"

Deleted: of

Deleted: in answer

Deleted: to

Deleted: , the search is coun

Deleted: ted as one

Deleted: provided of course

Deleted: in the gallery

Deleted: in the gallery.

Deleted: The motivation for this approach is twofold: if a gallery is returned after a search, the time required to work through the gallery is expected to be fairly stable, irrespective of the number of records in the gallery, and the time to process two galleries consisting of one record each is expected to be considerably more than the time required to process one gallery containing 2 reco

Deleted: rds.

Deleted: AAMVAUID9

Deleted: Draft 1.0

5.2.1.2. Relevance

False match rates are relevant in large-scale 1:N applications due to (1) the potential inconvenience they can pose to customers incorrectly flagged as matching another individual in the database and (2) the resources required for investigation and dispensation of false matches (above and beyond resources dedicated to resolving legitimate duplicate matches).

5.2.2. False Non-Match Rates

5.2.2.1. Definition

A biometric system's false non-match rate (FNMR) similarly bears different definitions depending on the type of system being evaluated. A universal definition of FNMR is

> *The probability that an individual's template will be incorrectly judged to not match that same individual's enrollment template.*

In other words, the false non-match rate represents the probability that a 1:1 comparison of two biometric records from the same individual will result in a non-match. For the purposes of 1:N systems, the false non-match rate indicates the percentage of individuals capable of successfully establishing a second[9] identity in a biometric system compared to the total number of multiple enrollment attempts.

$$\text{FNMR} = \frac{\text{\# of successful enrollments under a second identity}}{\text{total number of enrollment attempts to enroll under a multiple identity}}$$

5.2.2.2. Relevance

False non-match rates are relevant in large-scale 1:N applications inasmuch as they are strong indicators of the likelihood that an individual will be capable of establishing multiple identities within a database. Because the most visible objective of large-scale 1:N applications is to detect, deter, and prevent the enrollment of these multiple identities, maintaining acceptable false non-match rates is often a critical program objective, one

[9] For the purpose of simplicity, the Report focus is on a biometric system's ability to detect a second enrollment, although an individual may be capable of enrolling several times. The ability to enroll under a third or fourth identity is a function of the system's FNMR: because an individual attempting to establish a third identity will already have two biometric enrollments in the database (if one assumes that the user did not fail to enroll), the system only needs to match against one of the two previous enrollments in order for that person to be flagged and those "identities" removed. Note that it may be desirable, when an individual with duplicate enrollments is detected, to execute further 1:N searches at less restrictive thresholds in order to locate further fraudulent enrollments.

Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Deleted: <#>Page Break<#>False Non-Match Rates¶ <#>False non-match rates are defined in order to establish a frame of reference against which reported false non-match rates from biometric tests and deployments can be compared.¶
Deleted: ¶
Deleted: ¶
Deleted: imposter
Deleted: ¶
Deleted: imposter
Deleted: second
Formatted: Bullets and Numbering
Deleted: AAMVAUID9
Deleted: Draft 1.0

central to program credibility.

A major challenge in measuring false non-match rates in operational environments as opposed to test environments is that jurisdictions will most likely be unaware of the occurrence of false non-matches. Unless multiple identities are located by means other than the primary biometric technology or technology combination, false non-matches may not be resolved.

Deleted: has occurred

Formatted: Bullets and Numbering

5.2.3. FMR and FNMR in 1:N Systems

Reporting results from 1:N searches is complicated by the fact that a search which returns one incorrect record can be considered both a false non-match and a false match. Also, a search which returns two records, one of which is correct and one of which is incorrect, is both a correct match and a false match. Clarity on this matter can be gained by classifying 1:N performance according to *Open Set Identification Performance Metrics*.

Formatted: Bullets and Numbering

5.3. Open Set Identification Performance Metrics

An emerging approach to defining success and error rates in 1:N systems is increasingly discussed in U.S. and international standards committees*: Open Set Identification Performance Metrics.* This approach has emerged from the evaluation of biometrics in watchlist applications, in which only a very small percentage of individuals interacting with biometric systems are likely to be present in the gallery database. However, this approach to defining error rates is suitable to almost any 1:N application, including the 1:300m application at hand. In order to understand the value of this method of measuring 1:N biometric system performance, a brief discussion of closed set vs. open set identification is necessary.

Formatted: Bullets and Numbering

5.3.1. Closed Set Identification

Closed set identification applications are those in which the individual being identified is *known* to be in the database being searched. Examples of closed set identification could include an enrolled employee presenting himself at an access control point or an enrolled citizen in a voter ID database attempting to enroll under a different identity. In both cases, the individual is already in the database. If the user being identified is the best match returned from the 1:N search, the result is *correct user identified*. If someone other than the individual being identified is the best match returned from the 1:N search, the result is *incorrect user identified*. Closed set identification does not utilize a matching threshold, but

Deleted: AAMVAUID9

Deleted: Draft 1.0

instead is predicated on simply returning the closest match from a 1:N search. Closed set identification applications can also be configured to return candidate lists comprised of the top N matches from the closed set search.

Formatted: Bullets and Numbering

5.3.2. Open Set Identification

Open set identification applications are those in which the individual *may or may not be* in the database. Examples of open set identification may include an individual registering for a public benefits program who may or may not have already previous registered for benefits, as well as an individual presenting himself at a 1:N access control point who may or may not be identified. Because the individual presenting himself for identification may or may not be in the database, a "correct" response may be that no match is found that exceeds the required threshold – which is never the case in closed set identification. Therefore the range of potential outcomes from a 1:N search is broader than the "correct" and "incorrect" results from closed set identification. A critical element of open set identification is that a comparison between two or more records must exceed a specified matching threshold in order to be reported as a match.

The metrics that measure these potential outcomes of open set identification were comprehensively described in the U.S. Department of Defense's *Face Recognition at a Chokepoint* test[10] published in 2002. The discussion below generally mirrors this approach. However, the DoD test was dedicated to evaluating watchlist functionality (in which the gallery being searched is small, and enrollees are not added to the watchlist subsequent to enrollment). Metrics were described relative to their applicability in a watchlist environment. However, these metrics apply to more than just watchlist applications – they apply to all 1:N open set applications. Most real-world applications (as opposed to controlled tests) are effectively open-set applications: deployers have no way of knowing for certain whether or not any individual presenting himself for identification is already enrolled in a 1:N database.

The following are the potential outcomes of 1:N open set searches; for the purposes of discussion, we will assume that the application is detection of multiple enrollees in a DL/ID program.

[10] **Face Recognition at a Chokepoint - Scenario Evaluation Results**, 14 November 2002, DoD Counterdrug Technology Development Program Office.

Deleted: AAMVAUID9

Deleted: Draft 1.0

Enrollee in Database

1. **Individual In Database Identified.** In this scenario, an individual in the database (e.g., a multiple enrollee in a benefits issuance application) is correctly matched above a certain matching threshold against his prior enrollment. This is one of two potential correct outcomes of 1:N open set searches: a multiple enrollee has been located.
2. **Individual In Database Identified As Another Individual.** In this scenario an individual in the database (e.g., a multiple enrollee in a benefits application) is incorrectly matched above a certain matching threshold against a different person in the database. Not only is the correct person not located (enabling multiple enrollment fraud), but resources must be expended in investigating the incorrectly identified record.
3. **Individual In Database Not Identified.** In this scenario, a multiple enrollee is incorrectly not matched above a certain matching threshold against his or her prior enrollment. This would allow a person to establish multiple identities in a 1:N system. This error maps to traditional false non-match reporting, above.

Enrollee Not in Database

4. **Individual Not In Database Identified As Being In Database.** In this scenario, an individual not enrolled in the database is incorrectly matched above a certain matching threshold against another individual's enrollment. Resources must be expended in investigating the incorrectly identified record. This error maps to traditional false match reporting, above.
5. **Individual Not In Database Not Found In Database.** In this scenario, an individual not enrolled in the database is correctly not matched above a certain matching threshold against any another individual's enrollment. This is one of two potential correct outcomes of 1:N open set searches.

Figure 13: Open Set Identification Performance Metrics

Because in the large majority of actual 1:N deployments the enrollee is not in the database, and the vast majority of 1:N transactions in civil ID applications do not involve attempts to enroll multiple times, performance in 4 and 5 above especially critical. Open set identification applications can also be configured to return candidate lists comprised of all candidates whose matching score exceeds threshold N.

Ideally, large amounts of biometric performance data rendered in terms of Open Set Identification Performance Metrics would be available. However, aside from the aforementioned DoD test, evaluations whose results can be mapped to Open Set

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

Identification Performance Metrics are nonexistent. Therefore, such an approach cannot currently be relied on for decisions in the 1:300m environment.

5.4. Failure To Enroll Rates

A biometric system's failure to enroll rate (FTER) represents the probability that an individual will be unable to enroll in a biometric system. A failure to enroll occurs when a biometric system is unable to capture or extract data from one or more biometric samples sufficient to generate a reference template. The failure to enroll rate includes individuals who lack the specific characteristic required to enroll (e.g. missing one or more fingers or hands) as well as those whose characteristics are not sufficiently distinctive, stable, or measurable to enroll.

Failure to enroll rates are critical in large-scale systems, as failures to enroll may require alternate forms of multiple enrollment detection to ensure that individuals cannot easily establish multiple identities.

Note that database size does not have a direct impact on failure to enroll rates as it does with FMR and FNMR. With the exception of systems configured to increase sample quality requirements on large systems in order to reduce FNMR, many systems may have a static FTER regardless of database size.

5.5. Interrelation of Error Types

Biometric error types – FMR, FNMR, and FTER – are directly interrelated. Any given error rate, such as a particularly low FMR or FNMR, can only be evaluated in a meaningful fashion when the corresponding error rates are known.

Reducing false match rates, or increasing the percentage of multiple enrollees detected in 1:N searches, results in a corresponding increase in the percentage of first time enrollees incorrectly flagged. Reducing false non-match rates, or decreasing the percentage of first-time enrollees incorrectly flagged, results in a corresponding increase in the percentage of multiple enrollees not detected in 1:N searches.

In addition, a less widely known interrelation between false non-matching and failure to enroll impacts the evaluation of biometric systems. In order to reduce failure to enroll rates, a system administrator may adjust thresholds such that lower-quality biometric data is accepted to form enrollments. This lower-quality data may result in an enrollment that is more susceptible to false non-matching, should the individual attempt to re-enroll using another identity. This interrelation has not been studied as extensively as has the FMR-

Formatted: Bullets and Numbering

Deleted: <#>Failure to enroll rates are defined in order to establish a frame of reference against which reported failure to enroll rates from biometric tests and deployments can be compared. ¶ <#>Definition¶

Deleted: one or more biometric samples,

Deleted: to

Deleted: ,

Deleted: ¶ **<#>Relevance ¶**

Deleted: Failure to enroll rates are critical in large-scale systems, as failures to enroll result in the creation of a set of individuals capable of establishing multiple identities without fear of detection by biometric systems. ¶

Deleted: image

Deleted: (see below)

Deleted: FTE

Formatted: Bullets and Numbering

Deleted: FTE

Deleted: imposter

Deleted: non-imposters (

Deleted:)

Deleted: imposter

Deleted: AAMVAUID9

Deleted: Draft 1.0

FNMR interrelation.

An additional error type subsumed under FNMR and FTER is the failure to acquire rate. A failure to acquire occurs when a biometric system is unable to capture a biometric sample, or to extract biometric data from a biometric sample, sufficient to generate a reference template or match template. Therefore a false non-match or failure to enroll may be attributable to the inability of imaging technology to provide a usable biometric sample to the system's enrollment and matching algorithms. From a deployer perspective, the failure to acquire rate is important inasmuch as improved imaging technology (sensors or cameras) can reduce failure to acquire and provide a higher percentage of users' data to biometric algorithms.

Deleted: FTE

Deleted:

Deleted: ¶
¶
Page Break
Biometric Technologies Capable of 1:N Matching ¶
¶
Facial recognition, fingerprint, and iris recognition are the three technologies evaluated for suitability for a 1:300m application. ¶
Not included in this list of 1:N technologies are voice-scan (or speaker verification) and DNA verification. Voice-scan technology can be used for small-scale identification, on the order of dozens or hundreds of records, and has been leveraged for criminal forensics-style applications. However we are currently unaware of any deployments in which the technology has been deployed to databases of tens or hundreds of thousands of enrollees. In addition, the behavioral nature of voice-scan is such that attempts to mask or alter one's voice in a fashion not perceptible by 3rd parties may further undermine the ability to execute 1:N identification. As data emerges from studies or deployments in which voice-scan performs large-scale identification, the technology may warrant further evaluation. DNA is not yet developed as an automated identification technology in the biometric identification sense; manual processing is still necessary. It can also be argued that DNA identification is based less on measurement of a characteristic than on the material retention of a physical sample. However, DNA identification technology may eventually come to resemble biometric identification as currently implemented, such that the technology could be used to perform large-scale automated identification.¶
The following discussion of facial recognition, fingerprint, and iris recognition technologies' core operations, strengths, weaknesses, and maturity provides a framework for subsequent discussions of tests, deployments, and provider data.¶
¶
Page Break
Facial Recognition¶
Facial recognition technology is based on features such as the location and composition of distinctive features of the face, as well as ... [1]

Formatted: Bullets and Numbering

Deleted: AAMVAUID9

Deleted: Draft 1.0

6. Test Efforts

Understanding the three primary types of biometric testing, as well as their objectives, benefits, and limitations, is essential to determining if biometrics can be used effectively for 1:N matching. Much of the data relevant to determining if biometrics can be used for 1:N matching is generated through biometric tests. Vendor projections may not be accurate reflections of real-world performance; operational data, for its part, may be heavily dependent on a specific application's parameters. In addition, new types of tests may need to be executed in order to gauge technology capabilities.

Deleted: <#>Types of Biometric Testing¶

Deleted: should the Report reveal that insufficient data is available to determine the viability of 1:300m identification, biometrics tests of various types

Deleted: proposed and

Formatted: Bullets and Numbering

6.1. Test Types

6.1.1. Technology Testing

Technology testing involves comparison of biometric data through one or more matching algorithms. Databases used in technology testing may be collected specifically for the purpose of testing, or may be culled from operational systems. In either case, reduced emphasis is placed on biometric acquisition, as the data used for testing has already been acquired, normally through a single sensor. This approach is elaborated on in *Best Practices in Testing and Reporting Performance of Biometric Devices*[11]*:*

Deleted: offline

Deleted: a

Deleted: database

Deleted: against

Deleted: Such d

> *The goal of a technology evaluation is to compare competing algorithms from a single technology. Testing of all algorithms is carried out on a standardized database collected by a "universal" sensor…Testing is carried out using offline processing of the data. Because the database is fixed, the results of technology tests are repeatable.*

Deleted: ¶

Technology testing, despite its drawbacks, is the most relevant to determining the viability of successful 1:300m matching. Technology testing is more likely to be large-scale (encompassing 100k+ records) than other test types. In most cases a specific test application is written to enable 1:N testing and to generate results of "all vs. all" comparisons of biometric databases. Technology tests have been published on facial recognition, fingerprint, and iris recognition systems.

[11] http://www.cesg.gov.uk/site/ast/biometrics/media/Best%20Practice.pdf

Deleted: AAMVAUID9

Deleted: Draft 1.0

6.1.2. Scenario Testing

Scenario testing involves the collection and matching of biometric data from test subjects in a controlled test environment meant to emulate a specific biometric usage scenario. Scenario testing evaluates biometric sensors (cameras, scanners, microphones, etc.) in conjunction with their associated matching algorithms as full biometric systems. Scenario testing involves real-time, transactional comparison of an individual's biometric data against an enrollment acquired through a specific system. Scenario testing introduces multiple enrollees specifically for the purpose of attempting verification and/or identification as legitimate users. As such, scenario testing can often generate more extensive FMR data than operational testing.

Deleted: imposter

This approach is elaborated on in *Best Practices in Testing and Reporting Performance of Biometric Devices:*

> *The goal of scenario testing is to determine the overall system performance in a prototype or simulated application. Testing is carried out on a complete system in an environment that models a real-world target application of interest. Each tested system will have its own acquisition sensor and so will receive slightly different data. Test results will be repeatable only to the extent that the modeled scenario can be carefully controlled.*

Scenario testing is not generally applicable to determining the viability of successful 1:300m matching. Scenario testing generally utilizes small test populations, such that it is difficult to enroll and test sufficient users to generate results on large databases. Scenario tests have been published on facial recognition, fingerprint, and iris recognition systems.

6.1.3. Operational Testing

Operational testing involves measurement of the performance of biometric systems in an actual deployment with actual users. As opposed to a static database or a controlled test population, operational testing measures system performance in the field. This approach is elaborated on in *Best Practices in Testing and Reporting Performance of Biometric Devices:*

> *The goal of operational testing is to determine the performance of a complete biometric system in a specific application environment with a specific target*

Deleted: AAMVAUID9
Deleted: Draft 1.0

population. In general, operational test results will not be repeatable because of unknown and undocumented differences between operational environments.

Operational testing provides both benefits and drawbacks. The considerable variables that impact biometric performance are, by definition, incorporated within operational tests, as they are measurements of actual system operation. Assuming that the operational data being evaluated was collected in a biometric application similar to one's own, highly applicable data may be derived. However, it is difficult to precisely measure accuracy in operational tests. In a 1:N environment, the deployer is largely unaware as to whether a 1:N false non-match has taken place (i.e. the individual has gone undetected in the creation of multiple identities). Very little data is published from operational tests, primarily due to concerns regarding security and competitiveness.

Deleted: with precision

Formatted: Bullets and Numbering

6.1.4. Probes and Galleries

1:N biometric testing utilizes probes and galleries. A probe is comprised of X records used for the purposes of matching against Y enrollees. A gallery is comprised of Y enrollees against which X records are searched. In a DMV environment, the probe corresponds to the X new enrollees providing biometric data; the gallery corresponds to the database of Y enrollees against which the probe is searched.

Deleted: AAMVAUID9

Deleted: Draft 1.0

Test Type	Key Aspects	Advantages	Disadvantages	Relevance to 1:300m ID
Technology	• Offline testing • Data contained in single database compared against 1 or more algorithms • Can use legacy databases • Matching elements separated from acquisition process	• Can execute large-scale testing (100k+ records) due to elimination of acquisition process • Tests are repeatable • Can execute multiple offline tests to gauge performance with different controls (e.g. age, gender) • Enables head to head algorithm comparison	• Does not provide a strong measure of FTER – individuals tested may have already been vetted for enrollment • Minimizes impact of acquisition processes, which may skew results	• High; only test type that combines scalability with controlled depiction of match and non-match errors
Scenario	• Evaluates biometric sensors (cameras, scanners) along with their associated matching algorithms • Attempts to mirror a real-world application • Predicated on live interaction between a test subject and one or more biometric systems, as well as a dedicated test population	• Able to project FMR, FNMR, and FTER to a biometric application • Use of dedicated test subjects ensures consistent data quality, system interaction • Can control enrollment and matching effort	• Resource-intensive test processes preclude large-scale testing • Users may interact with systems in a test environment differently than in an operational environment • Can be difficult to accurately mirror a usage environment	• Low, due to lack of scalability
Operational	• Test of a fielded, deployed system *in situ* • Collects data from actual users conducting actual transactions	• No need to create test environment or emulate matching processes –uses real data • Operational issues impact test outputs	• Difficult to measure false non-matches in 1:N systems; such users avoid detection • Data reporting a challenge	• Moderate; however, FNMR is difficult to measure; slight changes in operating environment can skew data; results can be difficult to report

Figure 14: Biometric Test Types

Deleted: FTE
Deleted: S
Deleted: FTE
Deleted: Figure 11
Deleted: Draft 1.0

6.2. Facial Recognition Testing

Of the three technologies under consideration, more comprehensive test results have been published for facial recognition than any other technology. The availability of large databases from which to draw facial images, as well as the high degree of speculation and interest regarding the 1:N performance of facial recognition performance, enable and drive such facial recognition testing.

The results from these tests strongly suggest that facial recognition cannot successfully perform single-record identification on a database of 300m users, and that conducting facial recognition searches on a database of this size with a typical daily enrollment load would result in a preponderance of false matches, false non-matches, or both.

Although facial recognition has been evaluated since the mid 1990's through test efforts such as the FERET program[12] and Face Recognition Vendor Test (FRVT) 2000[13], the two most recent published efforts are **Face Recognition Vendor Test (FRVT) 2002**[14] and the U.S. Department of Defense's **Face Recognition at a Chokepoint**[15].

Deleted: Checkp

Deleted: [16]

6.2.1. FRVT 2002

6.2.1.1. Overview

The 2002 Face Recognition Vendor Test (FRVT 2002) represents a large technology test of the ability of facial recognition to perform 1:1, watchlist, and 1:N identification. The results from this test, by far the largest published facial recognition test (conducted with approximately 37,000 subjects), indicate that the best facial recognition systems encounter a false match rate of over 25% when executing 1:N searches on databases comprised of 37,000 enrollees.

Deleted:

FRVT 2002 involved two tests, the High Computational Intensity (HCI) test and the Medium Computational Intensity (MCI) test. HCI test data is more relevant to the question of 1:300m identification. 8 leading systems were tested in the HCI test, representing the state of the facial recognition art as existed in mid-2002. The fact that several systems were tested is important, as findings are not necessarily tied to any given technology or matching

[12] http://www.itl.nist.gov/iad/humanid/feret/feret_master.html
[13] http://www.frvt.org/FRVT2000/default.htm
[14] http://www.frvt.org/FRVT2002/Default.htm
[15] http://www.dodcounterdrug.com/facialrecognition/DLs/ChokePoint_Results.pdf
[22] http://bias.csr.unibo.it/fvc2002/

Deleted: Face Recognition at a Chokepoint - Scenario Evaluation Results, 14 November 2002, DoD Counterdrug Technology Development Program Office¶

Deleted: Draft 1.0

approach. The inclusion of multiple systems also demonstrated extreme differences in capabilities from system to system.

FRVT 2002 provides a good deal of useful information on the identification of multiple enrollees through 1:N searches (that is, the false non-match rate and true match rate for multiple enrollees). It increases our understanding of how well systems can locate multiple enrollees. The information generated regarding 1:N searches of legitimate users – expected to be by far the more common scenario – is less comprehensive. Most of the testing done was executed with the "new enrollee" already in the database, which is inconsistent with the 1:300m concept of operations laid out previously. Therefore we have less information on the false match rate and true non-match rate for legitimate first-time enrollees.

Deleted: imposter

Deleted: imposter

6.2.1.2. Population

Deleted: Test Population

The testing used a total of 121,589 images from 37,437 individuals (at least 3 images per individual). These images were acquired from a Department of State database comprised of photo images from visa applications. These images were acquired in a controlled, operational environment – but were not collected solely for the purposes of the test. Nearly all of the facial images were taken from the Department of State's Mexican non-immigrant visa archive. A small percentage of images were of individuals of Chinese descent.

6.2.1.3. Execution

Deleted: Test Execution

FRVT 2002 was based on comparison of facial images in a probe database (corresponding to "new enrollees" in a system) against a gallery database (corresponding to a "previous enrollee" database). The gallery database contained the earliest – first chronological – dated image from each of the 37,437 subjects. The probe database contained two additional images from each subject: the most recent image and the image closest to the temporal median between the oldest and newest images. Therefore if 300 days had elapsed between first and last image acquisition, the probe database would contain imaged acquired on day 150 (or thereabouts) and 300; the gallery contains the image from day 1. In 1:N testing, images in the probe database were searched against all images in the gallery database. With 1:1 and 1:N test loads combined, approximately 15 billion total matches were executed.

Deleted: N

Substantial time lapsed between the first and last facial image for many of the enrollees. The time between initial and subsequent image capture was reported in 60-day intervals up to 1140 days. Therefore this data may provide a reasonable reflection of identification

Deleted: Draft 1.0

performance up to three years after enrollment. Test results were also categorized by age in 5-year blocks from 18 to 78, in order to determine the degree to which age impact matching capabilities, as well as by gender.

This is a critical consideration for a 1:300m matching environment, as individuals will be enrolled over a period of years, and multiple enrollees bent on establishing multiple identities may return 1 day, 1 week, or 1 year after their initial enrollment. 1:300m capabilities may differ dramatically between 1 day and 1 year after enrollment.

Deleted: imposter

6.2.1.4. FRVT 2002 Results

FRVT results demonstrate the difficulty of executing 1:N searches on even modestly-sized databases. The following table indicates, for participating systems, the percentage of multiple enrollees who were incorrectly matched as another individual in the 37,437-person database. The best system incorrectly identified over ¼ of fraudulent enrollees as a different individual, such that the true enrollee would have been undetected. Performance seems to decline in a linear fashion as the database grows larger, such that on a database with between 10m and 100m enrollees, this error rate for the best system may increase to 60%. However, such testing has not yet been published.

Deleted: clearly

Deleted: imposter

Deleted: million

1:N FNMR for 37437-person DB	
System 1	30.0%
System 2	26.0%
System 3	35.0%
System 4	74.0%
System 5	60.0%
System 6	69.0%
System 7	72.0%
System 8	88.5%

Figure 15: False Non-Match Rate – Incorrect Match Rate for Duplicate Enrollee

It is also worth noting that on a 100-person database, the best-performing system still reported an FNMR of 9%.

All systems had lower FNMR for males, while ½ had lower FMR and ½ had higher FMR for males. Most systems recognize older people (50+) better than younger – for some systems, the difference was a few percentage points, for others error rates were four times lower for older individuals than for younger individuals.

Deleted: higher

Deleted: Draft 1.0

The following table demonstrates the considerable impact of time on 1:N accuracy. These results suggest that performance on a large database will vary dramatically according to the time elapsed from initial and subsequent biometric data acquisition. The best systems are able to correctly identify a multiple enrollee from the database over 80% of the time within 120 days of enrollment. After 120 days, the ability to correctly identify a multiple enrollee drops to the point where less than 70% are identified after 1 year and 60% after approximately 3 years – even for the best-performing system.

Deleted: ¶

1:N FNMR (%) for 37437-person DB from 0-3 years																			
Days	**0-60**	**60-120**	**120-180**	**180-240**	**240-300**	**300-360**	**360-420**	**420-480**	**480-540**	**540-600**	**600-660**	**660-720**	**720-780**	**780-840**	**840-900**	**900-960**	**960-1020**	**1020-1080**	**1080-1140**
System 1	20.0	23.0	25.5	28.0	30.0	31.0	32.0	31.0	34.0	34.0	35.0	35.5	35.5	36.0	33.0	38.0	38.0	45.0	44.0
System 2	**17.0**	**19.5**	**22.5**	**24.0**	**25.5**	**27.0**	**29.0**	**28.0**	**31.5**	**30.0**	**31.5**	**33.0**	**34.5**	**33.5**	**31.5**	**33.5**	**37.0**	**41.0**	**40.0**
System 3	24.0	27.0	29.0	33.5	34.0	35.0	38.0	37.5	40.0	40.0	42.0	41.5	45.0	44.0	43.5	45.5	50.0	52.0	52.0
System 4	63.0	67.5	70.0	72.0	74.0	75.0	77.0	77.0	78.5	78.5	79.5	78.5	80.0	81.0	80.0	81.0	81.5	85.0	84.0
System 5	48.0	51.5	55.0	57.0	58.5	60.5	63.0	62.0	65.0	64.5	67.0	65.0	68.0	68.0	67.5	67.5	70.0	72.0	68.5
System 6	59.5	63.0	65.0	67.5	70.0	70.0	72.0	71.0	72.5	72.5	75.0	73.0	74.0	74.5	74.5	76.5	78.0	79.5	78.0
System 7	61.5	66.5	67.5	70.0	72.5	72.0	72.5	73.0	77.0	76.0	77.5	75.0	78.0	78.0	78.0	78.0	80.0	83.5	80.0
System 8	83.0	86.0	86.5	88.5	88.5	89.5	90.0	90.0	91.0	91.0	92.0	90.5	91.0	92.5	92.5	92.5	93.0	92.5	90.5

Figure 16: Effect of Time on 1:N capabilities

Deleted: Bold = Best Performer

Deleted: Draft 1.0

6.2.1.5. Additional Commentary on Results

It is uncertain whether the ethnic composition of the database positively or negatively impacted the facial recognition technologies' ability to perform identification. There are two considerations here: first, that facial images of Mexican nationals may be easier or more difficult to match than those of different ethnic backgrounds, such that identification may be easier or more difficult; second, that a database with greater ethnic variation (Caucasian, African-American, Asian etc.) may be easier or more difficult to search than one of monolithic ethnic composition. Intuitively, it seems that searching a database comprised of multiple ethnic groups would generate more robust identification scores, as the physiological differentiation apparent across ethnicities should eliminate some percentage of prospective targets in a 1:N search. Further study in this area is necessary. For the purposes of 1:300m identification, the fact that certain ethnic groups may be susceptible to higher or lower rates of identification is noteworthy and demands investigation. If a certain biometric is capable of identifying 95% of duplicate enrollees across the full population, but identifies 90% of a certain ethnic group and 98% of another ethnic group, claims of bias may be introduced.

It is worth noting that FRVT 2002 did not distinguish between matching errors and enrollment/acquisition errors. Therefore it is uncertain what percentage of errors was attributable to acquisition and what percentage was attributable to matching. The test report comments accordingly:

> *If a system could not extract a template for a gallery signature, then most likely, all matches with that signature would produce low similarity scores. This would be reflected in the performance statistics. The FRVT 2002 design is transparent to these types of internal errors that may be handled differently by each participant system. …Failure to acquire is handled in the same manner as a failure to enroll.*

Individuals who posed errors across systems were not listed, so it is unclear whether matching algorithm fusion would have reduced error rates.

Deleted: Draft 1.0

6.2.1.6. FRVT Watchlist Testing

In addition to the 37,000+ database test detailed above, FRVT 2002 included watchlist testing of facial recognition systems. Watchlist testing is important to the question of 1:300m identification because it is consistent with the biometric concept of operations in which an individual *not previously enrolled in the system* is searched in a 1:N fashion against an enrollment database (or gallery). In fact, it is very safe to assume that nearly all 1:N transactions in a 1:300m DL/ID system will be executed in this fashion, as most enrollees will not be attempting to circumvent or defraud the system and will thus not be previously enrolled.

FRVT watchlist testing provides a "false alarm rate" for enrollees not present in the gallery and a "correct alarm rate" for individuals already present in the gallery. This testing was on a smaller scale than the primary FRVT 2002 testing: 3000-person gallery and probe databases were used. In addition, watchlist searches were conducted against smaller galleries of 50, 100, 200, 400, 800, 1600, and 3000 users. This testing differs from the primary FRVT testing in that images were acquired specifically for this testing, with users directly facing the camera and posing for image acquisition (Bold = Best Performer).

Rank 1 FNMR (%) with 1% FMR							
DB Size	**25**	**50**	**100**	**500**	**1000**	**2000**	**3000**
System 1	24.0	31.0	31.0	38.0	40.0	43.0	45.0
System 2	**23.5**	**25.0**	**26.5**	**37.0**	**39.5**	**42.0**	**44.0**
System 3	24.0	28.0	30.0	38.5	40.0	42.0	44.0
System 4	78.0	85.5	87.5	97.0	98.5	99.0	100.0
System 5	55.0	63.0	63.0	73.0	74.5	77.5	78.0
System 6	60.0	67.0	70.0	83.0	88.5	93.0	94.0
System 7	63.0	73.0	73.0	80.0	82.0	83.5	84.5
System 8	84.5	88.0	90.0	93.0	94.5	96.0	96.5

Figure 17: 1:N Testing for Multiple and Legitimate Enrollees

Deleted: Bold = Best Performer

Deleted: Draft 1.0

These results indicate that a system configured for a 1% false match rate (such that 1 of 100 first-time enrollees would be flagged), the best system fails to identify over 40% of duplicate enrollees on a data base of only 3000 individuals.

6.2.1.7. Applicability

The results from FRVT 2002 are likely to be reasonably applicable to those acquired in a DL/ID environment, although results may improve substantially based on use of multiple frontal images to create an enrollment template. In order to acquire multiple angles on enrollment, either stereoscopic cameras would need to be deployed during photo capture or the individual will need to provide data as requested by the enrollment agent.

These results strongly suggest that facial recognition is not accurately scalable to databases of hundreds of millions of individuals.

6.2.2. DoD Scenario Test – Face Recognition at a Chokepoint

6.2.2.1. Overview

The Department of Defense Scenario Test – Face Recognition at a Chokepoint was a test of a single biometric system's ability to perform 1:1 and watchlist identification. As above, watchlist testing is important to the question of 1:300m identification because it is consistent with the open set identification concept of operations in which an individual not previously enrolled in the system is searched in a 1:N fashion against an enrollment database (or gallery). The test utilized small databases of up to 1,575 users, against which 144 individuals attempted to match. If facial recognition were to struggle to perform adequately at this size, then its ability to match against a 300m-enrollee databases is suspect.

6.2.2.2. Methodology

The watchlist evaluation uses the scenario of attempting to find individuals at a chokepoint (i.e. metal detector). Users walking through the chokepoint stop and look at the system cameras for approximately three seconds, then continue walking. The system continuously compares found faces to images in a watchlist. This watchlist is made up of moderate-quality and high-quality facial images. The system displays the highest match that exceeds a certain threshold, along with a candidate list of other potential matches in descending order of similarity. The operator manually compares the top match with the live user and responds if an actual match is determined. While this concept of operations differs from photo image utilization, as comparisons are continual and not based on a single image, it is

Deleted: with
Deleted: Checkp
Deleted: [17]
Deleted: Checkp
Deleted: biometric
Deleted: perform against
Deleted: of millions of enrollees
Deleted: likely
Deleted: ¶
Deleted: Draft 1.0

likely that the number of images acquired as well as the image quality are such that results are reasonably extensible to a 1:N DL/ID application.

6.2.2.3. Results Reported

The performance data reported for this testing included the following:

- **Person on watchlist (POWL) correctly identified:** this is identical to the scenario in which a multiple enrollee is correctly matched against his or her prior enrollment.
- **POWL incorrectly identified:** this is identical to the scenario in which a multiple enrollee is incorrectly matched against another enrollee (i.e. false non-match rate).
- **POWL not alarmed:** this is identical to the scenario in which a multiple enrollee is incorrectly not matched against his or her enrollment (i.e. false non-match rate).
- **Non-POWL alarmed:** this is identical to the scenario in which a first-time enrollee is incorrectly matched against another individual's enrollment (i.e. false match rate).

Deleted: (

Deleted: (i

Formatted: Bullets and Numbering

Results are provided at varying security thresholds. The italicized column represents a non-error condition, i.e. correct identification rate.

Security	Person on Watch List Matched Incorrectly	Person on Watch List Not Matched	*Person on Watch List Matched Correctly*	Person Not on Watch List Matched Incorrectly
100 Targets – Low Quality Database Images				
High	19.2% (155/806)	62.8% (506/806)	*17.9% (145/806)*	29.4% (209/712)
Medium-High	3.6% (29/806)	85.7% (691/806)	*10.7% (86/806)*	4.1% (29/712)
Medium	1.7% (14/806)	90.8% (732/806)	*7.4% (60/806)*	0.4% (3/712)
Low-Medium	0.5% (4/806)	94.8% (764/806)	*4.7% (38/806)*	0%
Low	0%	96.7% (779/806)	*2.1% (17/806)*	0%
100 Targets – High-Quality Database Images				
High	8.7% (70/806)	15.1% (122/806)	*76.2% (614/806)*	38.1% (271/712)
Medium-High	2.7% (22/806)	34.1% (275/806)	*63.2% (509/806)*	9.1% (65/712)
Medium	**0.6% (5/806)**	**46.4% (374/806)**	***53.0% (427/806)***	**2.9% (21/712)**
Low-Medium	0.4% (3/806)	58.4% (471/806)	*41.2% (332/806)*	0.8% (6/712)
Low	0.2% (2/806)	66.0% (532/806)	*33.7% (272/806)*	0.3% (2/712)

Figure 18: 1:N Testing through Multiple Databases

Deleted: Draft 1.0

These results indicate that with 806 match attempts against a 100-person watchlist, a leading facial recognition system performs as follows (medium threshold):

- 47% of multiple enrollees were not identified or were identified incorrectly
- 53% of multiple enrollees were identified correctly
- Approximately 3% of legitimate first-time enrollees were incorrectly matched against an individual in the database.

Also, results indicate that image quality plays an extremely important role in this matching process.

Deleted: on

Deleted: database

Formatted: Bullets and Numbering

Deleted: Draft 1.0

6.3. Fingerprint Testing

In spite of the technology's broad deployment, few independent tests of fingerprint identification technologies have been published. Commonly discussed tests, such as the Fingerprint Verification Competition[22], focus on verification as opposed to identification. Pursuant to legislation passed subsequent to 9/11, large fingerprint tests have been designed and executed, with more large tests to follow.

6.3.1. NIST Standards For Biometric Accuracy, Tamper Resistance, Interoperability

6.3.1.1. Overview

The National Institute of Standards and Technology (NIST) has conducted large-scale tests on fingerprint and facial recognition technology; FRVT, above, was a NIST effort.

In mid-2002, NIST perform testing to define accuracy standards for 1:1 and 1:N biometric systems suitable for usage in entry/exit systems, addressing provisions of the PATRIOT Act and the Enhanced Border Security/Visa Reform Act. Results from this testing provide insight into 1:N fingerprint capabilities, but are only indirectly applicable to AAMVA's task. NIST's stated preconditions/assumptions in determining which technologies to test included the following:

- *The technology tested must be available and established.*
- *Large-scale (>100k) databases must be available for testing.* NIST implemented this requirement in order to ensure reasonable confidence in the statistical viability of the results.
- *Images for the biometric technology must be available.* This requirement precluded iris technology, as iris databases are mostly built on templates as opposed to images.
- *New and promising technologies can be tested and certified later.* This leaves room for iris and other emerging systems.

The following data relates to NIST testing of fingerprint technology executed in mid- to late 2002[24], referred to herein as the VTB testing. The VTB test represents a large-scale technology test of fingerprint verification and identification capabilities. We focus here only

[24] *NIST I R 7020 Studies of Fingerprint Matching Using the NIST Verification Test Bed (VTB),* http://www.itl.nist.gov/iad/894.03/NISTAPP_Nov02.pdf

Deleted: Iris Recognition Testing¶
<#>Of the three technologies under consideration, less independent testing has been conducted on iris recognition than any other. One of the major impediments to iris recognition testing in comparison to facial recognition and fingerprint is the lack of legacy databases against which to execute matching. However, the increased interest in facial recognition technology on the part of government agents for applications such as border control is likely to result in a vast increase in independent testing. ¶
<#>The perception of iris recognition as a highly accurate technology is based on results from internal tests, some of which have taken place on small controlled databases and others of which have taken place on larger operational databases. Iridian, the primary developer and patent-holder of iris recognition technology, has executed a number of internal tests of its technology in order to determine resistance to false matching. The largest and most recent test, entitled Iridian Cross Comparison Test, was published in December 2002. ¶
The following data must be read with the understanding that this testing was executed by the … [2]

Deleted: [18] and (20 a scenario test executed by the UK Centre for Mathematics and Scientific Computing National Physical Laboratory

Deleted: [19]. ¶
<#>Iridian Cross-Comparison Testing¶
<#>Overview¶
Iridian's Cross Comparison Test

Deleted: [20] represents a large technology test of the susceptibility of iris recognition to false matching when conducting 1:N matching. This test provides useful information for the … [3]

Deleted: [21], approximately 120,000 iris templates were collected from various operational databases from around the world. Several gallery databases were created from the following operationa … [4]

Deleted: As a result of

Formatted: Bullets and Numbering

Deleted: [23]

Deleted: ¶

Deleted: Draft 1.0

on identification.

6.3.1.2. Population

1:N testing took place against two fingerprint databases, each with over 600,000 subjects. The first database, referred to as DHS2, housed at least two left and two right index fingerprints from over 600,000 subjects. Fingerprints were flat impressions, not rolled[25]. Most images were acquired in operational border control environments from individuals crossing from Mexico into the U.S. The second database, referred to as DOS, also housed at least two left and two right flat index fingerprints from over 600,000 subjects, acquired in Mexican Consulates offices.

6.3.1.3. Execution

1:N testing utilized an open source fingerprint algorithm, in contrast to the other test efforts addressed in this Report, all of which used proprietary matching algorithms. The VTB test platform includes commercial off the shelf computer hardware, an open-source OS, and public domain application software.

Testing against the DHS2 and DOS databases followed a similar design. 1000 subjects were chosen at random from the gallery to execute matching against the 600,000-subejct database. Left and right index fingerprints were tested separately.

6.3.1.4. Results

The result returned, similar to FRVT 2002, was the percentage of 1:N searches that resulted in the correct person being identified from the database. Results are for a *single* fingerprint, not for a combination of both prints.

The results of the DHS2 testing indicated that on a database of approximately 600,000 users between 66% and 83% of individuals were correctly matched as the first match for left index fingerprints; between 70% and 86% were correctly matched as the first match for right index fingerprints. The variation within each fingerprint's results is explained by the fact that results were returned as averages for (10) groups of 100.

The results of the DOS testing, which used data acquired in a slightly more controlled operational environment. indicated that on a database of approximately 600,000 users, between 67% and 77% of individuals were correctly matched as the first match for left index

[25] It is uncertain to what degree rolled fingerprints would have improved performance relative to this test's flat fingerprint databases.

Deleted: Test Population
Formatted: Bullets and Numbering
Deleted: VTB
Deleted: environments
Deleted: patrol
Deleted: into the U.S. from Mexico
Deleted: .
Deleted: Test Execution
Formatted: Bullets and Numbering
Deleted: VTB
Formatted: Bullets and Numbering
Deleted: forgiving
Deleted: Draft 1.0

fingerprints; between 76% and 89% were correctly matched as the first match for right index fingerprints. The variation within each fingerprint's results is explained by the fact that results were returned as averages for (10) groups of 100.

On a database of 100,000 users, the identification rate on the DHS2 database was between 88% and 90% (left and right index). On a database of 100,000 users, the identification rate ranged widely from 76% to 86% (left and right index).

As reported by NIST[26]:

> *Single finger identification can provide 95% accuracy for a gallery size of 500. The identification rate drops to 90% for a gallery size of 10,000 and to 86% for a gallery size of 100,000. This test illustrates the difficult nature of accurate database searches using a single fingerprint. High accuracy searching of a database of 1 million subjects or greater will require more than one finger whether the FBI's IAFIS is used or not.*

Formatted: Bullets and Numbering

6.3.1.5. Applicability

The results of this data are relevant inasmuch as they demonstrate the need for multiple fingerprints to execute large-scale identification reliably, but little more than that can be derived. Within the biometric industry, there is an understanding that multiple-fingerprint systems can scale to over 1m users (dozens of such systems are deployed around the world).

Deleted: .

However, a methodological issue present in this testing limits the applicability of even this data to the 1:300m identification question. This testing provides useful information in answering the question, "will a multiple enrollee be the first match returned?" The possible answers are yes, the correct record was returned, or no, the wrong record was returned. This is acceptable for situations in which one is looking for a multiple enrollee. However, in almost all 1:N applications, most transactions involve a new enrollee who is *not* in the database. Therefore the more important question is whether the search returns a record above a certain match threshold X. This is because it is impossible to research every rank 1 match when the vast majority of enrollees are legitimate, first time users not already in the database. The biometric test community describes 1:N testing that utilizes thresholds as watchlist testing, but this test approach is necessary to gauge performance for all 1:N systems, not just watchlist. The DoD Chokepoint Test of facial recognition is the only 1:N

Deleted: above

[26] http://www.itl.nist.gov/iad/894.03/NISTAPP_Nov02.pdf

Deleted: Draft 1.0

test to correctly address this problem: the "watchlist" database is the same as the gallery database.

Deleted: ¶

Formatted: Bullets and Numbering

6.3.2. Philippines AFIS Benchmark Testing

Formatted: Bullets and Numbering

6.3.2.1. Overview

The U.S. National Biometric Test Center conducted testing[27] in the late 1990's to evaluate the ability of AFIS vendors to execute 1:N classification and matching on groups of fingerprints taken from test subjects. The testing was designed to facilitate decision-making on implementation of a 20m-person public benefits system in the Philippines and to develop a model for performance on what was, during that period, projected as an exceptionally large biometric database. Testing consisted of roughly 4000 fingerprints matched in a cross-comparative fashion to generate approximately 16m matches.

Formatted: Bullets and Numbering

6.3.2.2. Population

To create the gallery database (against which matches were executed), eight flat fingerprint images from thumb to ring finger were acquired from each of 510 Filipino office workers, roughly divided between male and female. This resulted in a gallery database with just over 4,000 unique fingerprints. This data was acquired under highly controlled and directed conditions. To create the probe database, the same eight fingerprints were acquired from 506 test subjects; 409 were also in the gallery (allowing for generation of FMR and FNMR). Biometric acquisition for the probe database was less strictly controlled, but still supervised.

Formatted: Bullets and Numbering

6.3.2.3. Execution

Participating AFIS vendors, of whom there were four, were required to first classify all fingerprints according to global fingerprint characteristics such as loops and arches. This allowed the testers to evaluate how effectively such classification schemes were in grouping fingerprints (a process which limits the number of records searched in a 1:N comparison). Vendors then cross-compared the two fingerprint databases in order to generate 1:1 match scores (which would then be mapped to a 1:N decision environment).

Formatted: Bullets and Numbering

6.3.2.4. Results

Matching results indicated that the best performers performed the probe vs. gallery

[27] The Philippine AFIS Benchmark Test Results, NATIONAL BIOMETRIC TEST CENTER COLLECTED WORKS 1997-2000

Deleted: Draft 1.0

matching with zero false matches at a single-finger FNMR of less than 10%. Worse-performing systems showed a single false match with a single-finger FNMR of 20%. Based on the overall data available, the test author concludes that a single-finger FMR of 1 per 1m records, with a corresponding FNMR of 10%, is representative of the strongest performers.

Classification results, important to limiting the number of records searched in a 1:N system, indicated single-finger penetration rates of approximately 50%. This suggests that before any 1:N matching commenced 50% of the database could be eliminated from the search process due to differences in the global fingerprint characteristic. For multiple-fingerprint systems, such classification enables searches against a fraction of the entire database.

By projecting single-finger match results to a multi-fingerprint system, and incorporating the penetration rates above (and also using gender-based filtering), the test projects that a two-fingerprint system can address deployer requirements for scalability to over 8m records. The multiplicative effects of single-finger FMR, penetration rates, and multiple fingers reduce the incidence of false match rates even in 1:N searches.

6.3.2.5. Applicability

While these results were generated several years ago, they do demonstrate the ability of systems using multiple flat fingerprints to scale to address large-scale database requirements.

6.3.3. Upcoming Test Effort: FpVTE

A recently announced fingerprint identification test effort, whose objectives, parameters, and methodology are similar to those of the Face Recognition Vendor Test 2002, should significantly expand the understanding of fingerprint identification capabilities. The Fingerprint Vendor Technology Evaluation[28] (FpVTE), also executed by NIST, will represent a technology test based on large database of operational fingerprint images, including rolled and flat images. The FpVTE will include a Large Scale Test (LST) that tests 50,000 tenprint fingerprint records, and as such may entail up to 2.5 billion comparisons.

Assuming that the recently disclosed test plan is retained, the critical element of FpVTE from a UID9 perspective is that it will represent the first published test to detail the performance of fingerprint systems that use multiple flat images. While general estimates as to the potential performance of tenprint rolled fingerprint systems can be offered, based on use in IAFIS and

[28] http://fpvte.nist.gov/index.html

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: Draft 1.0

similar criminal fingerprint systems, no such information is available on 6-, 8-, or 10-fingerprint flat image systems. Given the scale and operational constraints of the 1:300m identification application, a tenprint flat image system may be a requirement; however, testing is required to determine how accurate and scalable such a solution would be.

Depending on the timeframes that face UID9 decision-makers, access to test results – which may be available within the 2003 calendar year – may provide answers to critical performance questions and thus inform recommendations and decisions on technology standards.

Formatted: Bullets and Numbering

Deleted: Draft 1.0

6.4. Iris Recognition Testing

Of the three technologies under consideration, less independent testing has been conducted on iris recognition than any technology. One of the major impediments to iris recognition testing in comparison to facial recognition and fingerprint is the lack of legacy databases against which to execute matching. However, the increased interest in facial recognition technology on the part of government agents for applications such as border control is likely to result in a vast increase in independent testing.

The perception of iris recognition as a highly accurate technology is based on results from internal tests, some of which have taken place on small controlled databases and others of which have taken place on larger operational databases. Iridian, the primary developer and patent-holder of iris recognition technology, has executed a number of internal tests of its technology in order to determine resistance to false matching. The largest and most recent test, entitled Iridian Cross-Comparison Test, was published in December 2002.

The following data must be read with the understanding that this testing was executed by the vendor as opposed to an independent body. Independent published iris recognition tests include (1) an operational test executed against a very small (sub-1000) databases by the U.S. Department of Defense[29] and (2) a scenario test executed by the UK Centre for Mathematics and Scientific Computing National Physical Laboratory[30].

Formatted: Bullets and Numbering

6.4.1. Iridian Cross-Comparison Testing

6.4.1.1. Overview

Iridian's Cross-Comparison Test[31] represents a large technology test of the susceptibility of iris recognition to false matching when conducting 1:N matching. This test provides useful information for the evaluation of iris recognition technology for 1:300m identification. The results from this test, by far the largest published iris recognition test (conducted with operational data taken from approximately 120,000 subjects), indicate that iris recognition may be capable of performing identification against very large databases with a very low false match rate. As the HD between two iris templates grows closer to zero, more similarity is present between two templates. The decision policy, driven primarily by database size,

[29] *Testing Iris and Face Recognition in a Personnel Identification Application,* Dr. Steven King Information Systems Directorate Office of the Deputy Under Secretary of Defense
[30] http://www.cesg.gov.uk/site/ast/biometrics/media/Biometric%20Test%20Report%20pt1.pdf
[31] *Iridian Cross-Comparison Test,* December 2002, available for download at www.iridiantech.com

Deleted: *Cross Com*
Deleted: Draft 1.0

determines what HD is necessary to constitute a match. Templates with no similarity would have a HD of 0.50. Based on this decision policy, the effective FMR (according to vendor data) may be less than 1 in 2.79m for a single enrollee against a database on the order of magnitude of 100m records. This error rate is achieved by enforcing a decision policy whereby very low HDs are required for a match to be declared on searches of very large databases. Such performance is well within the bounds established for FMR in discussions with UID9. This is one of few tests, to our knowledge, that attempts to project performance to a level commensurate with the 300m person database under consideration.

However, this test leaves a number of areas unaddressed, and as such can only be judged "inconclusive" with regard to performance against a 300m person database.

Formatted: Bullets and Numbering

6.4.1.2. Population

In Iridian's testing, approximately 120,000 iris templates were collected from various operational databases from around the world. Several gallery databases were created from the following operational iris recognition databases (database sizes approximate):

Formatted: Bullets and Numbering

- 25,000 enrollees were from U.S. databases
- 20,000 enrollees were from Middle Eastern databases
- 65,000 enrollees were from South Asian databases
- 700 enrollees were from European databases

Iridian then created a 9000 person "probe" database comprised of iris data acquired from Icelandic deployments with which to search the target databases above.

Formatted: Bullets and Numbering

6.4.1.3. Execution

No duplicates enrollments were shared between the probe and the target databases, meaning that any matches that occurred could only be false matches. Large gallery databases were divided into databases of not more than 17,000 enrollees. The probe and target templates were matched as follows:

Formatted: Bullets and Numbering

- 9000 x 7374 = 66,366,000 matches
- 9000 x 16290= 146,610,000 matches
- 9000 x 17000 = 153,000,000 matches (several 17000-record databases were created)

This totaled 983m matches, a match load roughly equivalent to 1000 new enrollees being matched against a database with 1m enrollees. In the 983m matches executed, the following number of false match errors occurred at various thresholds.

Deleted: Draft 1.0

Cross-Comparison Testing generated only one metric: the number of false matches at a given threshold (or hamming distance). The Hamming Distance, or HD, represents the degree of similarity between two iris templates. As the HD between two templates grows closer to zero, more similarity is present between two templates. Templates with no similarity would have a HD of 0.50.

Formatted: Bullets and Numbering

6.4.1.4. Results

The results of the Cross-Comparison Test were as follows.

Formatted: Bullets and Numbering

- 157 errors (false matches) occurred at the 0.31 threshold (i.e. when the hamming distance between the two matched templates was 0.31 or lower)
- 32 errors (false matches) occurred at the 0.30 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.30 or lower)
- 10 errors (false matches) occurred at the 0.29 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.29 or lower)
- 3 errors (false matches) occurred at the 0.28 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.28 or lower)
- 1 error (false match) occurred at the 0.27 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.27 or lower)

Therefore, by enforcing a decision policy whereby an HD of less than 0.27 is required for a match to be declared, a 9000 person probe can be compared against a 100,000 person database with only one false match. No template comparisons resulted in an HD less than 0.26; it is uncertain how many more comparison would have been necessary for a false match to have occurred within this threshold.

Formatted: Bullets and Numbering

6.4.1.5. Areas not Addressed in Cross-Comparison Testing

A major gap in the relevance and extensibility of Cross-Comparison Testing data to a 1:300m matching environment is the lack of false non-match rate (or false reject rate) data. In order for a test's false match rate to be meaningful, it is necessary to determine the percentage of duplicate enrollees who would have evaded detection in a 1:N search. In order to generate such data, it is necessary to acquire at least two biometric records from a given individual (one for the probe, one for the gallery). As of the publication date of this report (December 2002), Iridian indicated that such results would be forthcoming.

Cross-Comparison Testing also does not address enrollment rates. The data used to execute testing was gathered from operational databases consisting of enrolled individuals.

Deleted: Draft 1.0

The lack of reported FTE may also impact the reported false match rates; if high enrollment quality thresholds were established for these databases, then no low-quality images would have been accepted for enrollment. This may or may not in turn impact false match or false non-match rates.

Formatted: Bullets and Numbering

6.4.1.6. Use of Single Iris

This test used for identification only a single iris from each individual. According to Iridian's internal test data, which must be validated or reproduced independently, an individual's left and right iris differ to nearly the same degree as different individuals' irises[32]. Therefore it is possible that utilizing both irises can provide a near-multiplicative effect on accuracy at scale (absent factors such as acquisition effort). This is in contrast to fingerprint technology, in which studies have shown that the physiology of fingerprints is such that the level of accuracy provided by N-fingerprint systems is much less than multiplicative. In other words, individuals' fingerprints are correlated, such that an individual with fingerprint characteristics *X* for a given finger has a greater than random chance than his or her other fingerprints will have characteristics similar to *X*.

Formatted: Bullets and Numbering

6.4.2. Iris Recognition Decision Policies

Because a single technology suppler is responsible for nearly all iris recognition deployments, the potential for iris recognition technology to be used for 1:300m identification is closely bound to the core technology of a single firm (Iridian). As opposed to most biometric providers whose systems are deployed with variable accuracy and security setting to account for differing deployer requirements, Iridian has historically held to a policy by which its systems are deployed with a false accept rate no greater than 1 in 1.2m. This rate applies to both 1:1 and 1:N systems, such that Iridian's technology will attempt to meet this 1 in 1.2m false *accept* rate (measured in terms of 1:N database searches) by driving its single-template false *match* rate (measured in terms of 1:1 matches) to extremely low levels. Regardless of database size or transaction loads, the threshold at which two templates are declared to match is automatically adjusted to enforce what Iridian has established through internal evaluation as a 1 in 1.2m false accept rate.

The reason that this policy is notable is that on a very large-scale database, Iridian's system would be designed to automatically drive false accept rates (what we refer to as effective false match rates) to a lower level than may be required in a 300m record system. As a

[32] http://www.cl.cam.ac.uk/users/jgd1000/genetics.html

Deleted: Draft 1.0

result, the system's false non-match rate (or false rejection rate, viewed from a 1:N perspective) may become excessively high. However, further testing is necessary to determine how the false non-match rate is impacted by such a decision policy. The following table from the *Iridian Cross-Comparison Test* demonstrates how Iridian enforces decision policies such that its false match rate remains low, regardless of database size.

HD Threshold	Database Size	Observed False Matches	Estimated False Match Rate	Estimated False Accept Rate[33]
0.31	10^0	157	1.60×10^{-7}	1.60×10^{-7}
0.30	10^1	32	3.25×10^{-8}	3.25×10^{-7}
0.29	10^2	10	1.02×10^{-8}	1.02×10^{-6}
0.28	10^3	3	3.05×10^{-9}	3.05×10^{-6}
0.27	10^4	1	1.02×10^{-9}	1.02×10^{-5}
0.26	10^5	0	3.0×10^{-11}	3.0×10^{-6}
0.25	10^6	0	3.16×10^{-12}	3.16×10^{-6}
0.24	10^7	0	3.16×10^{-13}	3.16×10^{-6}
0.23	10^8	0	2.79×10^{-14}	2.79×10^{-6}

Figure 19: Iris Recognition Decision Policy and False Match Rates

Formatted: Bullets and Numbering

6.5. Conclusions

While an increasing amount of biometric test results are available for analysis, challenges are present in drawing conclusions on performance against databases of hundreds of millions of users based on test data. First, there is no universally accepted method of testing biometric systems or reporting biometric test results. Each of the tests discussed above embraced different approaches to data collection, matching, and reporting. While NIST is planning development of a generic test suite to enable generic testing and reporting methodologies across technologies (and inclusive of multiple biometrics), such a framework has not yet been widely adopted. Therefore the data resulting from biometric tests is very much subject to interpretation, and is often heavily informed by the methodologies and assumptions of the test organizations. Second, biometric tests have taken place on much smaller databases than that under consideration, with even larger databases being limited to tens of thousands or hundreds of thousands. While performance at larger databases can be

[33] This is the equivalent of this Report's "effective false match rate".

Deleted: Draft 1.0

projected, there is no certainty that performance trends on smaller databases will continue through to very large databases.

These constraints notwithstanding, test results seem to demonstrate that iris recognition and fingerprint are theoretically capable of performing well against very large databases. With each technology, the ability to leverage multiple biometric characteristics from nearly all individuals is central to the scalability of iris and fingerprint technologies.

Facial recognition has been tested more thoroughly and independently than any other biometric; however, results for this technology suggest strongly that 1:N identification at the scale under consideration will result in unacceptable high error rates (well into the double digits for even the best performers).

While no testing has been executed on multimodal biometric systems, there is a broadly-held consensus among vendors and academics that combining biometric technologies in an intelligent fashion will improve overall performance. If it were determined that fingerprint and/or iris recognition were capable of operating within the performance bounds established by UID9, then a determination would need to be made as to the degree of improvement available through combinations of fingerprint, iris, and/or facial technology.

Deleted: ¶

Deleted: Draft 1.0

7. Deployments

7.1. Overview

Assessing the scale and type of deployments of facial recognition, fingerprint, and iris recognition technology provides additional information on the ability of these technologies to execute1:300m identification. In most cases performance data such as false match, false non-match, and failure to enroll rates from operational implementations is not available. This lack of available data may be attributable to an inability to measure all performance elements in an operational performance, to confidentiality agreements with technology providers, or to the risks of disclosing performance gaps that could enable system circumvention.

The lack of hard data notwithstanding, the scale, duration, and number of deployments associated with a given biometric technology provides useful information on the proven track record of a technology. Also, deployment information provides a real-world context for performance data gathered from biometric tests and performance projections on the part of technology providers.

At the outset it must be emphasized that no biometric deployments even begin to approach the scale of the 300m record system under consideration. The largest fingerprint deployments are in the tens of millions; the largest facial recognition deployments are in the millions; and the largest iris recognition deployments are in the hundreds of thousands. Of these technologies, fingerprint systems are most likely to have been deployed in systems whose concept of operations (1:N identification across an entire enrollment base) mirrors that of the 1:300m application.

Deleted: ¶
The composition of large-scale systems is expected to change dramatically over the next three years with the emergence of large-scale civil ID projects such as border entry/exit, national ID, and biometrically-enabled passports. Facial biometrics are to be incorporated in passports and machine-readable travel documents around the world, in accordance with ICAO recommendations and in compliance with U.S. legislative requirements. This will increase by orders of magnitude the number of individuals enrolled in facial recognition systems (although the processes by which 1:1 and 1:N matching for such systems is as yet undetermined). Iris recognition is currently positioned for deployment in border control projects in the Middle East, in South Asia, the U.K., and Canada. It is very likely that systems with well over 1m iris recognition enrollees will be populated within this timeframe. Therefore any decisions informed by the current lack of large-scale deployments of facial recognition and iris recognition technologies may be superceded by forthcoming events.¶
In addition, the adoption of multiple-fingerprint systems is expected to increase over the next three years, although the rate of adoption for fingerprint may not increase as rapidly as for facial and iris systems. ¶

Deleted: <#>NOTE: RESPONSES FROM TECHNOLOGY PROVDERS REGARDING CURRENT DEPLOYMENTS HAVE BEEN RECEIVED IN PART. WE ANTICIPATE RESPONSES, FULL OR PARTIAL, FROM SEVERAL ADDITIONAL PARTIES. THIS ADDITIONAL MATERIAL WILL BE PRESENTED AND ANALYZED ONCE SUFFICIENTLY REPRESENTATIVE RESPONSES ARE GATHERED FOR EACH TECHNOLOGY. … [5]

Formatted: Bullets and Numbering

Deleted: Draft 1.0

7.2. Facial Recognition

Facial recognition technology is deployed with 1:N capabilities in a handful of deployments with over 1m enrollees. Such 1:N systems are designed to return candidate lists for the purposes of facilitating fraud investigations, as opposed to returning only single records. Projects indicative of the scale at which the technology is deployed include the following:

- Illinois DMV: >13m records (*projected* 25m records)
- Colorado DMV: 10m records (including stored, multiple records per DL/ID holder)
- Australia Passports: 3.5m records

These large-scale facial identification deployments leverage existing imaging processes such as digital facial image capture, piggybacking on current processes to reduce the impact of biometric system implementation on current operations. Note that this differs from the optimal method of enrolling facial images by which an individual provides facial images from several different angles in order to constitute an enrollment template. Searches may return one or more matches for new enrollee, at which point the face image pairs are manually investigated.

7.3. Fingerprint

Fingerprint technology is by far the most commonly deployed biometric technology for identification, used in both criminal and civil ID systems. Projects indicative of the scale at which the technology is deployed include the following:

- Nigeria National ID: 60m records (6 fingerprints, flat placement)
- FBI IAFIS: 55m records (10 fingerprints, rolled)
- Philippines Social Security: 6m records (*projected* 35m records) (4 fingerprints, flat)
- Los Angeles County Sheriff's Department: *projected* 20m records (10 fingerprints, rolled)
- IDENT: 5m records (2 index fingerprints, flat)
- Malaysia Government Card: 18m records (2 thumbprints, flat)
- California Department of Justice: 12m-14m records (10 fingerprints, rolled)

Flat placement systems are more relevant to the 1:300m application under consideration, as addressed below. The following assessment of select fingerprint deployments provides insight as to their applicability to 1:300m DL/ID applications. It is important to note that the concepts of operations for 1:N fingerprint systems can differ substantially in terms of number of fingerprints acquired (1-10), rolled or flat acquisition, immediate or non-immediate

Formatted: Bullets and Numbering

Deleted: <#>¶
Page Break
<#>Facial Recognition¶
<#>Facial recognition technology is deployed with 1:N capabilities in a handful of deployments with over 1m enrollees. Such 1:N systems are designed to return candidate lists for the purposes of facilitating fraud investigations, as opposed to returning only single records. ¶

Formatted: Bullets and Numbering

Deleted: Draft 1.0

turnaround time, and number of placements required. This is in contrast to 1:N facial recognition and iris systems, each of which is implemented in a relatively consistent fashion. 1:N facial recognition systems normally generate candidate lists and use a single image for enrollment, while Iris recognition systems provide near-immediate single-record identification using the best-of-several enrollment images.

Deleted: <#>¶
I

7.3.1. AFIS

7.3.1.1. Scale and Scope

The largest biometric system, IAFIS, contains 55m records, each with 10 rolled and 10 flat fingerprint images, collected and converted over decades. IAFIS is used to conduct background check and criminal checks, and consists of both civil records (for employment background checks) and criminal records taken from booked suspects. IAFIS is capable of handling over 80,000 1:N searches per day, although this figure may grow higher as the current system's capacity is enhanced. This figure is well below the anticipated 130,000 enrollments per day required to populate an AAMVA-wide DL/ID system.

Deleted: 80m

Deleted:

IAFIS requires collection of ten rolled and flat fingerprints in order to enable large-scale searches and to enable search of partial latent prints (as required in criminal systems). Collecting this amount of data can be time-intensive, taking up to several attempts over a period of minutes for certain individuals. Specific image quality standards are required for a record to be searchable against the IAFIS database.

7.3.1.2. Relevance

IAFIS demonstrates the operational scalability of fingerprint technology to systems with tens of millions of records, searched with tens of thousands of new records per day. Given that the system was designed and deployed in several phases from the mid- to late 1990's, it is logical to conclude that improvements in AFIS technology since that time could be leveraged to enhance such a system's operations. At the same time, IAFIS provides a model for a potential nationwide biometric implementation inasmuch as it was the result of a concerted, multi-year effort that leveraged extensive standards development, technology improvement, and government funding to accomplish a specific set of objectives.

However, several aspects of IAFIS limit the relevance of this deployment to a broader-use, 300m record system. Although the current task is to evaluate fingerprint systems for potential utility in a 300m record system, a distinction must be drawn between criminal and

Deleted: Draft 1.0

civil fingerprint systems. The former are based on acquisition of rolled fingerprints, which requires either ink and cards or specialized optical scanning equipment; the latter utilize fingerprint placements, which provide much less usable data but are much easier to acquire. It is reasonable to view rolled and flat fingerprint systems as representing two different biometric technologies. Rolled fingerprints provide, in most cases, at least twice the usable data that flat placements provide, and as such essentially represent a different set of physiological characteristics than a flat fingerprint placement (flat fingerprint data is a subset of rolled fingerprint data). Being a criminal ID system, IAFIS is predicated on collection of extensive amounts of biometric data from both cooperative and non-cooperative individuals.

Therefore, while IAFIS demonstrates the scalability of fingerprint technology in systems with several tens of millions of records, this scalability applies to rolled fingerprint systems as opposed to flat systems. Understanding that the current task pertains to theoretical biometric technology capabilities, and that issues around viability of biometric acquisition in an operational environment are addressed subsequently, it is nonetheless important to recognize the fact that the current manner in which rolled fingerprints are acquired is completely inconsistent with mainstream DMV operations and would pose extraordinary logistical (as well as privacy and user perception) challenges[34]. The degree to which IAFIS performance could be extended to a flat fingerprint system – which would be much more in line with use in a DMV environment – is uncertain.

The problem of interoperability between flat and rolled fingerprints is currently under investigation as the IAFIS database and IDENT database (see below) are tested for compatibility. Initial estimates are that the use of four or more flat fingerprints to search a tenprint rolled database would provide "acceptable" performance, such that there is some potential for multiple flat print combinations to be used on applications at scale.

Deleted: leverage

Deleted:

Deleted: 4

Formatted: Bullets and Numbering

7.3.2. Nigeria National ID

7.3.2.1. Scale and Scope

The Nigerian National ID project involved the enrollment of over 50m individuals through 60,000 workstations over the course of three weeks. Six fingerprints were acquired from each enrollee. In order to enroll this quantity of individuals in this timeframe, train-the-trainer approaches to enrollment staffing were utilized. In this process, individuals provided each

[34] Efforts are underway to enable collection of rolled fingerprint images through hardware that eliminates the need to physically roll the fingerprint across a surface. The maturation of such equipment would greatly simplify the acquisition process and would make deployment in a civil ID system much more viable.

Deleted: Draft 1.0

fingerprint twice in order to confirm that the acquisition was useful.

7.3.2.2. Relevance

The Nigeria National ID project is relevant to the problem of 1:300m identification in a DL/ID application inasmuch as it is the largest application of biometric technology in a civil ID program. Although performance data is not yet available from this recent deployment, should it be made available, it would provide direct insight as to the performance of multiple flat fingerprints against large databases. As an indicator of scale, the acquisition of six fingerprints suggests that at least six fingerprints would need to be acquired for a 1:300m application (although it is unclear how many of the six fingerprints are being used for identification in the current system).

7.3.3. IDENT

7.3.3.1. Scale and Scope

The U.S. IDENT program is used at the U.S. Mexican border for real-time identification based on the acquisition of two flat fingerprints. Currently approximately 5m records are on files, searched with a sub-two minute response time.

7.3.3.2. Relevance

IDENT's concept of operations is more consistent with use in a DL/ID implementation than is IAFIS due to its collection of flat, as opposed to rolled, fingerprints. In addition, the system is designed for rapid turnaround (an average turnaround of 12 seconds, according to the vendor) from 1:N searches based on inputs from approximately 1000 workstations at distributed locations.

7.4. Iris Recognition

Iris recognition's largest deployments have recently reached the hundreds of thousands, representing a major expansion over previous projects. As iris recognition is adopted in the public sector in ID and border entry/exit programs, it is anticipated that even larger deployments will result. Nearly all iris recognition systems are deployed in a 1:N fashion, as is consistent with the vendor's preferred concept of operations. Projects indicative of the scale at which the technology is deployed include the following:

- Pakistan Humanitarian Assistance: 65,000 records
- United Nations High Commissioner for Refugees (UNHCR), Pakistan: 120,000 records

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: the

Deleted: 2

Deleted: -

Formatted: Bullets and Numbering

Deleted: <#>¶
<#>ADDITIONAL PERFORMANCE DATA FORTHCOMING¶

Deleted: . Deployments representative of the largest iris recognition implementations include the following:

Deleted: , approximately

Deleted: enrollees

Deleted: (ongoing).

Deleted: , approximately

Deleted:

Deleted: enrollees

Deleted: Draft 1.0

(*Projected* 1m over next 3 years).

- United Arab Emirates (UAE) border crossing: 200,000 records (ongoing) (*Projected* 1m over next 3 years).

In the first two applications, individuals are searched against an iris database to determine if they have already been issued an assistance package (in order to detect individuals attempting to collecting multiple such packages). These applications are predicated on the acquisition of a single iris from each enrollee.

Due to the relatively small size and lack of verifiable performance data, these implementations are not sufficiently large to draw conclusions on 1:300m performance. However, it is anticipated that emerging deployments in the U.K. and Canada could contribute substantially to an understanding of this technology's scalability. Early-stage national ID programs that may utilize iris recognition programs could also quickly result in deployments of over 1m users.

Decisions on deployment of iris recognition in large-scale applications are informed less by the history of large-scale deployments than by the demonstration of resistance to false matching through test result analysis. Also, Iridian has published results from an internal scalability test using 1m iris templates.

7.5. Multimodal Deployments

There are currently very few multimodal deployments in any biometric applications, and biometrics have not yet been deployed for multimodal identification in civil ID systems. Therefore no information is available on the performance or scalability of such a system. However, multimodal 1:N deployments are likely to emerge due to the use of biometrics in large-scale mandatory programs with various international stakeholders such as border entry and exit applications.

The International Civil Aviation Organization (ICAO) has identified facial recognition as the standard for biometrics in machine-readable travel documents and passports. ICAO's report also grants the option of using one or two secondary biometrics (fingerprint and iris recognition) to supplement facial recognition for personal identification.

Because the US VISIT program is driven by legislation which defers to ICAO technology recommendations for use of biometrics in machine-readable passports, ICAO's certification of facial recognition technology becomes critically important. If interpreted literally, U.S.

Deleted: *ongoing*

Deleted:). The project is expected to reach

Deleted: million

Deleted: enrollees over the

Deleted: .

Deleted: , approximately

Deleted: enrollees

Deleted: . The project is expected to reach 1 million enrollees over the next 3 years.

Deleted: The concept of operations of these 1:N implementations is such that it is difficult to determine when a user is falsely non-matched. In addition, the Pakistani deployments seem to be designed such that all matches, whether true or false, are treated as imposter attempts (no other information is collected from the individual). Therefore it is possible that what appear to be attempts at illegitimate reenrollment are actually legitimate first-time enrollees. Iridian reports a 0.6% FTE rate for the UNHCR deployment. ¶

Deleted: D

Formatted: Bullets and Numbering

Deleted: ¶

Page Break

Deleted: MRTDs

Deleted: Draft 1.0

legislation states that Visa Waiver Program countries need to incorporate facial recognition in their passports (in accordance with ICAO recommendations) in order to retain Visa Waiver Program status. Therefore countries such as the U.K. investigating the use of iris recognition for passport issuance would need to move to incorporate facial recognition in addition to iris recognition. Whether one or both of these two technologies would need to be utilized for identity verification is an issue unaddressed by the ICAO standard.

In terms of US VISIT (border entry/exit program) technology recommendations for visas, the National Institute of Standards and Technology (NIST) testing has generated recommendations that two fingerprints and one face image – stored in image form, most likely on a contact card – would provide sufficient 1:1 capabilities to warrant deployment for U.S. visa issuance applications. There is considerable controversy as to whether two fingerprints will be adequate to perform either internal 1:N searches or 1:N background check searches against the U.S. IAFIS database. In addition, the question of alignment with ICAO image standards has yet to be resolved. Also, the means by which these technologies will be used in conjunction have not been established, nor is there substantial research behind the combination of such technologies for transactional verification.

Deleted: , which contains 40m criminal and 40m civil ten-print records

The composition of large-scale systems is expected to change dramatically over the next three years with the emergence of large-scale civil ID projects such as border entry/exit, national ID, and biometrically-enabled passports. Facial biometrics are to be incorporated in passports and machine-readable travel documents around the world, in accordance with ICAO recommendations and in compliance with U.S. legislative requirements. This will increase by orders of magnitude the number of individuals enrolled in facial recognition systems (although the processes by which 1:1 and 1:N matching for such systems is as yet undetermined). Iris recognition is currently positioned for deployment in border control projects in the Middle East, in South Asia, the U.K., and Canada. It is very likely that systems with well over 1m iris recognition enrollees will be populated within this timeframe. Therefore any decisions informed by the current lack of large-scale deployments of facial recognition and iris recognition technologies may be superceded by forthcoming events.

In addition, the adoption of multiple-fingerprint systems is expected to increase over the next three years, although the rate of adoption for fingerprint may not increase as rapidly as for iris and facial systems.

Formatted: Bullets and Numbering

Deleted: Draft 1.0

7.6. Biometric Performance in Public Sector Identification Programs

Several of the larger applications of 1:N biometric technology have taken place in public sector ID programs, such as those in Arizona, New York, California, and Texas. While none of these programs have attained an enrollment base that even begins to approach that of the 300m required in this evaluation, any information available on performance would serve as a valuable indicator of real-world matching capabilities for systems with over 1m enrollees. It is worth noting that the traction that these systems gained in the mid- to late-1990's has slowed, such that very few new systems have been employed in the entitlements space.

Formatted: Bullets and Numbering

7.6.1. Difficulty of Measuring Performance

By definition, measuring performance in an operational environment is complicated by the fact that multiple enrollees may go unrecognized. A system may catch all multiple enrollees or may catch only a small fraction; in either case the deployer will most likely never know the system FNMR. Operational FNMR can be estimated by intentionally populating one's database with a number of test records – ideally sharing the same demographic breakdown and familiarity with the technology as real-world enrollees – and to attempt to enroll the individuals associated with these test records multiple times. Such testing would also need to take place over time; it is fair to assume that detecting a multiple enrollee the day after his or her enrollment will be much easier than such detection several months later.

While it is possible to measure enrollment errors and false match errors in operational systems, there is little published data available on such performance in operational environments. The biometric measurement of interest in all such programs is the number of multiple enrollees located. There is little perceived benefit to the deployer or to the vendor to disclose the percentage of users who cannot be enrolled or the percentage of times that the system errs, requiring manual resolution.

Other measurements often used in public benefits programs, such as cost avoidance attributable to multiple enrollee identification and to the deterrence effect associated with system deployment, are only tangentially related to actual matching and enrollment performance and are of minimal value in this discussion.

In most cases, requirements for accuracy are presented to vendors in procurement

Deleted: Draft 1.0

documents. For example, Texas' Lone Star Imaging System (LSIS)35 placed a requirement for 99% accuracy on its system. It is unknown whether this figure referred to both false non-match rates and false match rates, or simply false non-match rates (such that 99% of multiple enrollees would be detected). In addition, it is uncertain as to what type of testing was executed prior to award and subsequent to deployment.

Deleted: ¶

Formatted: Bullets and Numbering

7.6.2. Case Study: Colorado DMV

The use of facial recognition for 1:N multiple enrollee identification and subsequent fraud investigation on the part of the Colorado DMV provides some indication of performance in large-scale systems. Discussions with the Motor Vehicle Business Group Investigations Unit, who emphasized the utility and effectiveness of the system in fraud investigations, provide the following performance-related information.

A task force comprised of law enforcement, banking, and retail interests provided recommendations to the DMV regarding steps to be taken to reduce the incidence and impact of identity-related fraud. One of these recommendations was to incorporate biometric technology within its DL/ID issuance process.

Colorado DMV implemented facial recognition within its fraud investigation department in October 2002. Biometric system implementation entailed population of a central template database, with conversion of the entire image database, as well as deployment of a web-based application through which facial recognition functionality is accessible. Upgraded local PCs were required to operate the software. At the completion of each business day, every image acquired from new DL/ID applicants – approximately 3000 photo images – are searched in batch mode against the entire DL/ID photo database of approximately 10m images. The 10m images correspond to approximately 3.5m unique individuals, each of whom may have several images associated with their records. Of the approximately 3000 1:N searches executed on a daily basis, approximately 100-125 return one or more gallery images to be investigated. The current system is configured to return no more than 15 matches per individual; in most cases only a few matches – between 3 and 6 – are returned. Gallery images are returned based on operator-adjustable match thresholds.

Searches take place only on new applications: renewals are also not yet searched in a 1:N fashion, and there was no "all vs. all" search when the system was first implemented. The state's renewal process is currently 10 years; individuals only need to be imaged every other

[35] http://www.dhs.state.tx.us/providers/LoneStar/LSIS/

Deleted: Draft 1.0

renewal cycle, such that 20 years may pass between biometric data acquisition.

Investigators review the results of facial recognition searches, inspecting and marking potential those match which require further inspection. The “closer inspection” sequence entails the inspector viewing larger facial images, looking at signatures, and viewing thumbprint images (captured during the facial imaging process but not used for automated matching). If two images are determined to be a match, the records are flagged for further follow-up activity, most often requesting that the individual return to the DMV. Approximately 26 fraud cases are identified per month, or just over one per business day.

Though no data is available on the average template age of correctly matched imposter images in the gallery, operators report that in many cases searches correctly return multiple enrollment images acquired 3-4 years prior the new enrollment, suggesting that some level of multiple enrollment detection is attainable after substantial time.

The preceding data suggests that the system is configured such that approximately 3-4% of searches result in a match; of the 100-125 daily matches that comprise this 3-4%, approximately 99% are false matches and approximately 1% are actual matches. A strict calculation of FMR would also incorporate the fact that many of the 100-125 records provide multiple matches, such that an even higher percentage of matches are “false”. Mapping these results to the open set identification performance metrics provides the following information, which demonstrates the difficulty of drawing definitive performance conclusions even from systems whose results are shared.

1. Individual In Database Identified: percentage unknown; incidents approximately 1 per day.
2. Individual In Database Identified As Another Individual: percentage unknown due to lack of knowledge regarding database composition.
3. Individual In Database Not Identified: percentage unknown due to lack of knowledge regarding database composition.
4. Individual Not In Database Identified As Being In Database: approximately 3-4% of new enrollees were identified as being in the database, or approximately 100-125 incidents per day. It is reasonable to assume that since approximately 99% of these incidents were false matches, and that several records were returned with most searches for potential investigation, that in most cases the individual was not actually in the database.

Formatted: Bullets and Numbering

Deleted: Draft 1.0

5. Individual Not In Database Not Found In Database: percentage unknown due to lack of knowledge regarding database composition.

8. Performance Data from Technology Providers

Deleted: ¶
<#>¶
Page Break
<#>Performance Data from Technology Providers ¶
<#>

8.1. Introduction

IBG developed and issued questionnaires in June 2003 to gather information from biometric developers, vendors, and integrators on issues central to the execution of 1:300m matching. These questionnaires were accompanied by a cover letter explaining the auspices under which they were being distributed. Participation was at the discretion of the vendor, although recipients were encouraged to ensure that the data collected was representative of the state of the art in biometric systems technology and deployments.

Identical questionnaires were distributed to fingerprint, facial recognition, and iris recognition vendors. Recipients of the questionnaire were selected due to their involvement in the development and/or deployment of biometric technologies used for 1:N identification as well as their perceived capability to deliver 1:N biometric systems. To the degree possible, the responses synopsized and analyzed below have been genericized to avoid favoring any particular technology or vendor.

Deleted: reputation
Deleted: for being capab
Deleted: le of
Deleted: delivering
Deleted: for deployment

Performance data from technology providers who chose to address the problem of 1:300m identification (a subset of vendors surveyed) indicates that certain fingerprint, iris, and facial recognition providers see their technology as being capable of addressing the problem of 1:300m identification to varying degrees.

Deleted: .

8.2. Facial Recognition

Information provided by facial recognition vendors demonstrates the large-scale uses of the technology, while at the same time suggesting that accuracy on very large-scale database is an area not fully addressed on current-generation systems. Notable areas addressed included the following:

Formatted: Bullets and Numbering

- The one facial recognition vendor who chose to address scalability to 300m records estimated that an identification rate of 69% was attainable against such a database. This projection was based on results from on third party tests, including FRVT 2002 (assessed above), mapped to the projected performance of new algorithms projected to be available in 2004.
- The potential for multimodal biometric implementation in which facial recognition is used for an initial 1:N search and fingerprint technology is used for identity confirmation was

Deleted: Draft 1.0

emphasized. In this way a strength of facial recognition, rapid 1:N searches, can be leveraged to generate small candidate lists against which 1:1 fingerprint-based matching occurs.

- The importance of acquiring images in a fashion compliant with the emerging ICAO standard for facial imaging in machine readable travel documents was underscored. This acquisition methodology provides a baseline for image quality that ensures sufficient resolution for 1:N operations and ensures interoperability across all acquired images. Non-adherence to such image standards would results in reduced 1:N accuracy.

Deleted: <#>.¶

Reponses provided to the survey underscore the positioning of facial recognition as a technology capable of narrowing large databases to manageable sizes in 1:N applications as opposed to locating single records.

It is worth noting that the information provided by facial recognition vendors applies only to 2D systems as opposed to 3D systems. Nearly all facial recognition systems are 2D systems, particularly in the area of civil ID application. However, firms are aggressively researching 3D facial recognition as a solution to the problems of acquisition at an angle and of changes in facial appearance. With the exception of limited and non-applicable data published in FRVT 2002 on "morphable" facial recognition (in which 2D images are processed to assume the characteristics of 3D images), no information is available on the performance of such solutions.

Deleted: ¶

8.3. Fingerprint

Information provided by fingerprint vendors focused heavily on their experiences in large-scale system deployment, with deployments functioning as proof of scalability, viability, performance, etc. Of the vendors who chose to respond with projections of performance against a 300m person database, an emphasis was placed on the highly provisional aspects of such projections. Notable areas addressed included the following:

- As with facial recognition respondents, the potential for (and perhaps requirement for) multimodal biometric implementation was emphasized. Iris was cited as a technology that may provide strong synergies due to its high accuracy.

Formatted: Bullets and Numbering

- One respondent provided extensive information on the subfunctions and processes involved in executing 1:300m identification. This narrative, which encompasses the distinct stages of image quality analysis and rejection, fingerprint classification, minutiae extraction and comparison, and thresholding, seems to provide two important lessons.

Formatted: Bullets and Numbering

Deleted: Draft 1.0

The first lesson is that the number of variables which inform 1:N performance begins to mushroom once one isolates the separate components of the 1:N process. Each of the components involved entails utilization of separate technology components, performance measures, and filters to effectively move to the next stage. Also, as these components have been developed and enhanced over time using actual fingerprints gathered through sensors in different environments, they are fundamentally designed to account for the variability present in fingerprint acquisition, imaging, and matching processes.

To develop a conceptual evaluation model that isolates each factor impacting each component would be exceptionally difficult, given that (1) the performance of the components is interrelated and (2) each vendor's approach to solving the problem will vary. The second lesson is that approaches to identification at the 1:300m scale under consideration have been developed, and that the concept of 1:300m identification is not seen as overwhelming or illusory.

- In terms of actual performance projections, estimates based on internal testing against a database of approximately 100,000 records project that 4-6 flat fingerprints are capable of identifying 99.5% of duplicate records in a 300m database (acquisition of all ten fingerprints is recommended). As is the case with nearly all reported performance data from vendors or test centers, the methods by which this figure is derived and reported is unique to the vendor.

8.4. Iris Recognition

The performance data provided by iris recognition respondent(s) maps closely to the data discussed under Iris Recognition Testing, above, with the addition of estimated false non-match rates and failure to enroll rates (which were not provided in the published test results). Estimates based on internal testing indicate that a single iris can provide an identification rate of 98-99% (i.e. a FNMR of 1-2%) while falsely matching as few as 1 in 6m enrollees. These projections are based on enforcing a decision policy whereby very "low" match scores, or HDs, are required in 1:N searches.

8.5. Additional Response-Related Data

Reponses gathered from vendors were interesting for the information provided, the information volunteered but not requested, and for the information not provided. While the number of firms capable of effective 1:N identification in the field is limited, high-level

Formatted: Bullets and Numbering

Deleted: <#>International organizations that have had to issue identity documents (United Nations, OSCE, etc.) have studied the ratio between the average transactional capability and the transactional capability necessary to accommodate peak periods. The conclusions of these studies demonstrate that for the first issue of identity documents, the peak transactional capability is 2.5 to 3 times the average processing capability. For renewals and replacements, the peak throughput is only 20 to 30% higher than the average. The conclusion here is that the system can be sized for 120% to 130% of average steady-state throughput if new, biometrically credentialed licenses are issued at the normal enrollment time, but a system throughput of 250% to 300% of steady-state must be supported if an accelerated rollout of biometrically credentialed licenses is adopted. This will have a significant impact on cost. Through the rest of this document, we have assumed that the normal renewal cycles will be maintained, and that a 120 to 130% overdesign is sufficient to support the rollout period. ¶
<#>¶
<#>Depending on the frequency of the peak throughputs, customers have different approaches. Some have selected to size their system according to the anticipated peak transactional volume. This approach is typically adopted when peaks are expected every day (e.g., 11:00 a.m. to 3:00 p.m.). Alternatively, the system can be sized to accommodate the average volume and accept longer processing times during peak periods. Along with the database size and the transactional throughput, response time is a major AFIS sizing consideration. With a few exceptions, the largest AFIS are configured to process a specific number of searches per day, rather than to return a response within a certain period of time. ¶
<#>¶

Formatted: Bullets and Numbering

Deleted: Draft 1.0

observations on the vendor responses are as follows. While many vendors responded to the entire questionnaire, the assessment below applies primarily to vendor responses to questions that impact the Phase I assessment, including accuracy and deployment-related questions.

Vendors who elaborated on the question of 1:300m identification (as opposed to providing primarily numerical answers) invariably stressed the provisional nature of their responses, seemingly to underscore the difficulty of addressing these questions absent a full system specification for guidance. Vendors also provided, for the most part, unsolicited discussions on the practical issues involved in a 300m record system, including the importance of image quality, accounting for peak transaction times, and the unique constraints present in civil ID applications. Almost all vendors viewed the 1:300m question not as an abstract, theoretical exercise, but instead underscored how real-world constraints very heavily inform large-scale identification. For certain vendors, the contemplation of 1:300m identification absent the real-world considerations that impact any biometric application was seen as highly unrealistic.

Two respondents provided extensive discussion of the methodology by which 1:300m performance was projected, while a third cited combinations of internal projections and external tests. Vendors tended to project performance on larger databases by extrapolating the trends in the "correct identification" rate – the likelihood that a multiple enrollee will be the first person identified in a 1:N search – from smaller database sizes. For example, if the correct identification rate consistently drops by 3% with every order of magnitude increase in database size (from 100 to 1000 to 10,000 and so forth), then a vendor may extrapolate performance to 100m based on extension of this trendline. These projections leveraged internal database testing as well as results from third party evaluations.

It should be noted that certain vendors who refrained from providing performance projections in 1:300m applications cautioned against the use of extrapolated test data to project performance on larger databases. Other vendors noted the divergence in performance projected under controlled conditions, in which one can optimize performance to suit the characteristics of a target database, as opposed to performance in operational databases where less control is available over the data.

Vendors were forthcoming with information on their deployments, pointing to large databases and effective deployment processes as substantiation of their ability to address the challenges of 1:300m identification. There was considerable divergence of opinion in the

Deleted: Draft 1.0

area of the utility of multimodal biometric in executing 1:300m identification. Most respondents stated that multimodal biometrics would improve accuracy and enrollment rates, assuming that the logic through which multiple biometric solutions were combined addressed each technology's strengths and weaknesses. This logic would need to be inclusive of trends across different populations, of the relative weights of strong and weak matches and non-matches, and of the size of the database being searched. However, one respondent questioned the value of combining multiple biometrics, suggesting that the use of multiple biometrics will detract from overall system accuracy while also impacting throughput.

Deleted: Draft 1.0

9. Conclusions and Recommendations

Deleted: : Assessing the Viability of 1:300m Identification

Based on the current state of the art as reflected in published biometric tests, biometric deployments, and information gathered from biometric technology providers, insufficient data is available to determine conclusively whether any single biometric technology is capable of successfully performing 1:N identification on a database of 300m individuals within the bounds established through consultations with UID9.

9.1. Limitations of Available Information

Deleted: Gaps in

Deleted: Understanding

The following are the key gaps and limitations in the information at hand to determine how well biometrics may perform on databases of 300m records.

Lack of large-scale deployments from which to extrapolate performance. The IAFIS database consists of approximately 80m records, of which approximately ½ are civil records and ½ criminal records. The largest facial recognition databases number in the single millions, deployed against civil databases such as voter registration and DL/ID databases. The largest iris recognition deployments are in the hundreds of thousands. In most cases performance data such as false match, false non-match, and failure to enroll rates from operational implementations is not available. This lack of available data may be attributable to an inability to measure all performance elements in an operational performance, to confidentiality agreements with technology providers, or to the risks of disclosing performance gaps that could enable system circumvention.

Lack of Applicable Biometric Tests. While biometric testing remains the primary means by which one can gain controlled, objective test data on biometric system performance, drawing conclusions on performance in 1:300m applications is complicated by the following.

Deleted: it is very difficult to extrapolate biometric

Deleted: test

Deleted: to

Deleted: for several reasons

- *Number of Tests.* Only a small number of tests have been executed and published for each of the three primary 1:N technologies. Even if any test data were at hand to make determinations on 1:300m performance, such testing would need to be replicable to be of use in decision-making. No results have been published on multimodal 1:N testing.
- *Number of subjects.* Projecting performance on 300m-person databases is complicated by the lack of testing of systems with millions or tens of millions of records. Of the tests published, the largest iris test incorporated just over 100,000 subjects; the largest facial recognition test incorporated approximately 37,000 subjects; the largest fingerprint test incorporated approximately 600,000 subjects. Although one can test “all vs. all” to

Deleted: Draft 1.0

emulate the execution of billions of matches – as would be required daily in a 300m person system – this approach is limited by the fact that the test population is not comprised of independent samples, but instead a smaller number of (probe) samples are used to search the database. The results from searching one record against a 300m person database can differ from the results of searching 3000 records against a 100,000 person database, although the same number of matches will have been executed.

- *Methodology.* Most 1:N testing follows a methodology which generates results inconsistent with the concept of operations of most DL/ID transactions. The most central consideration in a 1:N test is whether the search returns a record above a certain match threshold X. Most testing focuses on the first record returned, as opposed to the score.
- *Testing over time.* While FRVT 2002 presents granular results of performance over time, which can have a direct impact on false non-match rates, most biometric testing either fails to report results over time or groups results into excessively broad categories. In a 1:300m environment, the time lapse between enrollment and subsequent identification can have a major impact on performance.

9.2. Evidence at Hand to Facilitate Decisions

The primary types of evidence that could be used to determine technologies' ability to scale to 1:300m identification are testing, deployments, and provider-based data. As of this writing, with information forthcoming from vendors on their deployments and internal data, it is too soon to determine the degree to which this data will be useful. Initial assessment of the utility of these three data types are as follows.

Evidence	Strengths	Weaknesses
Testing	• Enables objective performance measurement when executed independently • Most can be repeated with newer technologies to gauge improvement	• Methodology, process can limit relevance to real world • Difficult to acquire all necessary metrics in one test • Lack of tests
Deployments	• Operational constraints mirrored in system performance	• Performance can be impossible to measure • Little motivation to discuss or reveal operational performance
Provider-Based Data	• Providers should have clearest understanding of limits of technology	• Potential for bias, unrealistic assumptions in projections

Figure 20: Evidence Types

Deleted: Decisioning

Deleted: Initial assessments of the utility of these three data types is

Deleted: Type

Deleted: <#>Difficulty of data collection limits extensibility of data¶

Deleted: have

Deleted: technology

Deleted: conceptual

Deleted: Draft 1.0

9.3. Initial Evaluation of Technology Capabilities

Despite the lack of data indicative of 1:300m performance, provisional comments can be made regarding various technologies' capabilities in executing 1:300m matching.

Facial Recognition. Results from very small-scale tests indicate that facial recognition will not be capable of successfully performing 1:300m identification. The technology is subject to errors in which a substantial percentage of multiple enrollees are not located in databases of hundreds or thousands of users. More information is being gathered from vendors and deployers to determine if these tests misrepresent by several orders of magnitude the operational performance of facial recognition technology. Test results are not generally inclusive of FTER, such that evidence supplied by vendors and deployers will need to be gathered. Exceptionally high enrollment levels can be enforced through the use of manual processing in the event of a failure to enroll; as opposed to fingerprint and iris cases in which no such characteristic is present, one can project that nearly all of humanity can be enrolled in facial recognition systems given operator involvement.

Fingerprint. Based on experience with rolled fingerprint systems' scalability (to the several tens of millions of records), multiple flat-fingerprint solutions, particularly those which acquire 6, 8, or 10 fingerprint images from each enrollee, may be capable of identification on the order of 100m+ records. The acquisition of multiple fingerprints increases the quantity of distinctive data associated with a given enrollee. To date, little is known about the scalability of such solutions; we are unaware of any biometric system in operation based on ten flat fingerprint images, and testing of such systems is its incipient stages.

Iris Recognition. Testing internal to Iridian indicates that iris recognition decision policies can be enforced such that the single-record false match rate is kept exceptionally low, on the order of 2.79×10^{-14}. Such a decision policy would allow the effective false match rate (measured in terms of the probability of an individual matching against a 1:N database) to less than one incident per million, much lower than UID9 guidelines are likely to recommend. The relevance of this data is uncertain due to the lack of corresponding false non-match data, without which the above false match data cannot be reasonably assessed. Iridian has claimed FTER s of 0.6% in published work, but the definition of FTE can impact the actual FTER . In theory, aside from rare medical conditions and eye injuries, all individuals should be capable of enrolling in the technology, as the iris cannot be degraded as can a fingerprint.

Deleted: imposter

Deleted: rates

Deleted: .

Deleted: Iris Recognition. Testing internal to Iridian indicates that iris recognition decision policies can be enforced such that the single-record false match rate is kept exceptionally low, on the order of 2.79×10^{-14}. Such a decision policy would allow the effective false match rate (measured in terms of the probability of an individual matching against a 1:N database) to less than one incident per million, much lower than UID9 guidelines are likely to recommend. The relevance of this data is uncertain due to the lack of corresponding false non-match data, with out which the above false match data cannot be reasonably assessed. Iridian has claimed FTE rates of 0.6% in published work, but the definition of FTE can impact the actual FTE rate. In theory, aside from rare medical conditions and eye injuries, all individuals should be capable of enrolling in the technology, as the iris cannot be degraded as can a fingerprint. ¶

Deleted: s

Deleted: Draft 1.0

9.4. Focus Areas for Further Evaluation

Deleted: Areas for Further Research

Formatted: Bullets and Numbering

9.4.1. Development and Execution of Applicable Tests

While a number of biometric tests have been executed to gauge performance in various 1:N applications, these tests vary widely in terms of data collection approaches, matching and decision logic, reporting methodology, scale, duration – essentially in every parameter imaginable. A systematic and structured approach to evaluating biometrics in a fashion commensurate with the application under consideration is necessary. As demonstrated in this Report, the ability to compare and contrast biometric test results is limited. Generating data relevant to UID9 requires that two problems be solved: one of scale, the other of design. A large number of test subjects must be involved in any testing to begin to project performance on a 300m person database. However, just as important as test size is the structuring of a test approach that mirrors acquisition and instruction processes as found in a DMV environment and that executes 1:N matching in a fashion commensurate with DMV requirements.

Among the areas that warrant evaluation, due to the lack of published results, are the following:

- Performance over time (weeks/months/years)
- Performance across different populations (age/race/gender)
- Performance across different acquisition devices (high quality/high value)

Formatted: Bullets and Numbering

A particular area of interest is iris recognition performance in independent tests. Vendor-executed testing has demonstrated considerable promise in the ability to limit FMR on large-scale searches.

Formatted: Bullets and Numbering

9.4.2. Algorithm Fusion

Understanding the degree to which fusion of multiple algorithms can improve performance may be essential to developing highly scalable and accurate systems. Fusion biometric systems are those which process a single biometric input through multiple processing and matching algorithms to generate an enhanced result.

Formatted: Bullets and Numbering

9.4.3. Evaluation on Standardized Data Sets

Biometric systems can either leverage proprietary extraction, matching, and data formats, or can utilize standardized approaches. The large majority of tests and deployments

Deleted: Draft 1.0

referenced in this Report (with the exception of NIST fingerprint testing and the IAFIS fingerprint database) represent use of closed or proprietary technology approaches. However, it may be necessary to perform 1:300m identification through standardized formats, processes, and systems so as to avoid defining a proprietary approach that favors one vendor.

However, there is next to no data available on how well biometric systems perform when using standardized data formats or matching processes. It will be necessary to study the degree to which performance degrades (if any) when moving from closed to open biometric systems. It is also necessary to evaluate methods by which both closed and open approaches to matching can be leveraged for 1:N identification, as was done during the development of IAFIS in the 1990's.

Formatted: Bullets and Numbering

9.4.4. Multimodal Systems

Deleted: Until r

Deleted: is

Rigorous study must be conducted that generates accuracy rates for multimodal systems for comparison against strong single-biometric systems. Many vendors expressed that multimodal systems were a potential answer to the 1:300m challenge. However, very little data is available to demonstrate real-world improvements derived by multimodal systems. Testing across a broad population, to determine whether certain biometric characteristics are correlated to the point where multimodal operations are excessively impacted, is necessary.

Deleted: s, the question as to whether combining a "strong" biometric technology (one with low FMR and FNMR) and a "weak" biometric technology (one with high FMR and/or FNMR) provides better results than a single strong biometric will not be addressed.

Deleted: Draft 1.0

Appendix A: Large-Scale Testing Rollup

1:N Test / Biometric	Persons / Samples	Probe / Gallery	Composition	Aging	FMR[36]	FNMR	FTE	Notes
Completed Tests								
FRVT *(Facial Recognition Vendor Test)* 2002 HCI FACE	37,437 / 121,589	37,437 vs. 37,437	Visa applications (primarily Mexican nationals)	Up to 1140 days	Wrong target returned in 1:N search of 37,437: 17% within 60 days 27% within 1 year		Not calculated	• This FMR-FNMR for best of 8 systems tested • This corresponds to a DL/ID scenario in which a multiple enrollee attempts to enroll a 2nd time
FRVT *(Facial Recognition Vendor Test)* 2002 WATCHLIST FACE	3000 / 6000	6000 vs. 3000	Visa applications (primarily Mexican nationals)	Not reported – likely followed aging above	Enrollee not in 3000 person database falsely matched against db : 1%	Enrollee in 3000 person database incorrectly not returned as rank 1 match: 38%	Not calculated	• This FMR-FNMR for best of 8 systems tested • FMR corresponds to a DL/ID scenario in which legitimate user attempts to enroll for first time • FNMR corresponds to an AAMVA scenario in which a multiple enrollee attempts to enroll a 2nd time

[36] Due to the variance in results reporting, FMR, FNMR, and FTER may not be comparable from top to bottom.

Formatted: Bullets and Numbering
Deleted: "..." ... [6]
Formatted: Bullets and Numbering
Deleted: db ...4...- ...(span:... [7]
Deleted: provided herein for systems only ... [8]
Deleted: n...AAMVA ...n impo ... [9]
Deleted: Target not returned in db search of 34,437: **17% - 27%** (span: 1 year)
Deleted: IDENTIFICATION
Deleted: Probe ...fails to ... ... [10]
Formatted: Bullets and Numbering
Deleted: <#>FMR-FNMR provided herein for best systems only ¶ o an AAMVA ... ... [11]
Deleted: returns rank 1 ...- ... [12]
Deleted: WATCHLIST
Deleted: n...imposter ... [13]
Deleted: Draft 1.0

1:N Test / Biometric	Persons / Samples	Probe / Gallery	Composition	Aging	FMR[37]	FNMR	FTE	Notes
Completed Tests								
IRIDIAN CROSS-COMPARISON IRIS	120,000 / 120,000	9000 vs. 109304 (latter broken into 9 smaller databases)	Global (70% Middle East, 20% U.S., 10% other, no Africa/ South America)	n/a	Between 1 in 1.60×10^7, 1 in 1.02×10^9 depending on db size and threshold	Not calculated	Not calculated	• Total # of false matches increased (1, 3, 10, 32, 157) as threshold increased • FNMR not calculated – did not have image pairs • FTER not calculated, used previously enrolled subjects
NIST FINGERPRINT	620,000 / 3,000,000 (2-30 index prints/ subject)	1000 vs. 600,000	INS recidivists (primarily Mexican nationals)	Not reported – most likely substantial	Wrong target returned in search of 600,000: **11%-34%** of 100,000: **10%-24%**		**0.5-5.0%** (varies for rolled and flat databases)	• Used generic FP algorithms, equivalent of commercial technology from late 1990's • This corresponds to an AAMVA scenario in which a multiple enrollee attempts to enroll a 2nd time
Philippines SSO FINGERPRINT	607 / 8208 (thumb-ring for each subject)	4128 vs. 4080	Filipino office workers, 55% female	1-6 weeks, but temporal results not reported	**Single-finger FMR: 1.0×10^{-6}** Note: these results were mapped to a multi-print system	**2%-10%**	**1.0-2.5%**	• 409 of the 506 probe enrollees were included in the gallery • These figures are estimates – results were not reported in the test in a granular fashion

[37] Due to the variance in results reporting, FMR, FNMR, and FTER may not be comparable from top to bottom.

Formatted: Bullets and Numbering
Formatted: Bullets and Numbering
Deleted: (0.31): 1.5972×10^{-7} ¶ (0.30): 3.2553×10^{-8}¶ (0.29): 1.0173×10^{-8}¶ (0.28): 3.0519×10^{-9}¶ (0.27): 1.0173×10^{-9}
Deleted: m
Deleted: e
Deleted: the
Deleted: US
Deleted: Asia/
Deleted: FTE
Deleted: 450
Deleted: 5
Deleted: certainly
Deleted: INS, IAFIS civ, IAFIS crim)
Deleted: a few years old
Deleted: 4
Deleted: of 10,000: **10%** of 1,000: **7%**
Deleted: No target returned in search of 100,000: **14%** of 10,000: **10%** of 1,000: **7%**
Deleted: n
Deleted: imposter
Deleted: <#>Single finger penetration rate ¶
Deleted: rough
Deleted: guides – did not provide detailed outputs
Deleted: Draft 1.0

1:N Test / Biometric	Persons / Samples	Probe / Gallery	Composition	Aging	FMR[38]	FNMR	FTE	Notes
Pending Tests								
FpVTE *(Fingerprint Vendor Technology Evaluation)* 2003 FINGERPRINT	60,000 / 400,000 (various combinations of 1-10 fingerprints)	Multiple combinations	Operational US Federal Gov't databases	TBD	**TBD**	**TBD**	**TBD**	• Being executed 1 October 2003 through 21 November 2003

Figure 21: Large-Scale Biometric Test Efforts

[38] Due to the variance in results reporting, FMR, FNMR, and FTER may not be comparable from top to bottom.

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: Draft 1.0

Appendix B: Fundamental Biometric Concepts and Processes

Biometric systems convert data derived from behavioral or physiological characteristics into templates, which are used for subsequent matching. This is a multi-stage process whose stages, as well as associated technologies, are described below.

Verification, also referred to as 1:1 matching, identity confirmation or authentication, is the process of establishing the validity of a claimed identity by comparing a match template against a reference template. Verification requires that an identity be claimed, after which the individual's enrollment template is located and compared with the verification template. The result of a verification attempt is a score, which indicates the probability that the person is whom they claim to be. Verification answers the question, "Am I who I claim to be?"

Identification, also referred to as 1:N or one-to-many matching, is the process of determining a person's identity by searching a database of biometric templates. Identification systems are designed to determine identity based solely on biometric information.

There are two types of identification systems: positive identification and negative identification. Positive identification systems are designed to find a match for a user's biometric information in a database of biometric information. Positive identification answers the "Who am I?" although the response is not necessarily a name – it could be an employee ID or another unique identifier. Negative identification systems search databases in the same fashion, comparing one template (or perhaps several in the case of an automated fingerprint identification system) against many, but are designed to ensure that a person is not present in a database. This prevents people from enrolling twice in a system, and is often used in large-scale public benefits programs in which a person with bad intent might attempt to enroll multiple times in order to gain benefits under different names.

Enrollment is the process whereby a user's initial biometric sample or samples are collected, assessed, processed, and stored for ongoing use in a biometric system. Enrollment takes place in both 1:1 and 1:N systems. If users are experiencing problems with a biometric system, they may need to re-enroll to gather higher quality data.

Biometric samples are the identifiable, unprocessed image or recording of a physiological or behavioral characteristic, acquired during submission, used to generate biometric

Deleted: matching,
Deleted: , or identification
Deleted: ,
Formatted: Bullets and Numbering
Deleted: Draft 1.0

templates for enrollment and matching. The following sample types are associated with each biometric technology:

Technology	Acquisition Device
Fingerprint	Desktop peripheral, PCMCIA card, mouse, chip or reader embedded in keyboard
Facial recognition	Video camera, PC camera, single-image camera
Iris recognition	Infrared-enabled video camera, PC camera

Figure 22: Biometric Acquisition Device Types

Acquisition devices, also referred to as readers or scanners, are the hardware used to acquire biometric samples. The following acquisition devices are associated with each biometric technology:

Technology	Biometric Sample
Fingerprint	Fingerprint image
Facial Recognition	Facial Image
Iris Recognition	Iris Image

Figure 23: Biometric Sample Types

Feature extraction is the automated process of locating and encoding distinctive characteristics from a biometric sample in order to generate a template. The feature extraction process may include various degrees of image or sample processing in order to locate a sufficient amount of accurate data. For example, voice-scan technologies can filter out certain frequencies and patterns, and fingerprint technologies can thin the ridges present in a fingerprint image to the width of a single pixel. Furthermore, if the sample provided is inadequate to perform feature extraction, the biometric system will generally instruct the user to provide another sample, often with some type of advice or feedback.

The manner in which biometric systems extract features is generally considered proprietary, and varies from vendor to vendor. Common physiological and behavioral characteristics used in feature extraction include the following:

Technology	Biometric Characteristics
Fingerprint	Location, direction, and relative position of friction ridge endings and bifurcations on fingerprint; ridge line patterns
Facial Recognition	Relative position and shape of nose, position of cheekbones
Iris Recognition	Furrows and striations in iris

Figure 24: Biometric Characteristic Types

Formatted: Bullets and Numbering

Deleted: Draft 1.0

A **template** is a comparatively small but highly distinctive file containing data derived from the features of a user's biometric sample or samples. Templates are used to perform biometric matches. A template is created after a biometric algorithm locates features in a biometric sample. The concept of the template is one of biometric technology's defining elements, although not all biometric systems use templates to perform biometric matching: some voice-scan systems utilize the original sample to perform a comparison.

Depending on the purpose for which they are generated, templates can be referred to as reference templates (or enrollment templates) or match templates. Reference templates are normally created upon the user's initial interaction with a biometric system, and are stored for usage in future biometric comparisons. Match templates are generated during subsequent verification or identification attempts, compared to the stored template, and generally discarded after the comparison. Multiple samples may be used to generate a reference template – facial recognition, for example, will utilize several facial images to generate an enrollment template. Match templates are normally derived from a single sample – a template derived from a single facial image can be compared to the enrollment template to determine the degree of similarity.

The manner in which information is structured and stored in the template is generally proprietary to biometric vendors. Biometric templates are not interoperable – for instance, a template captured by one vendor's fingerprint system generally cannot be matched against a template generated in another vendor's system.

Different biometric templates are generated every time a user interacts with a biometric system. As an example, two immediately successive placements of a finger on a biometric device generate entirely different templates. These templates, when processed by a vendor's algorithm, are recognizable as being from the same person, but are not identical. In theory, a user could place the same finger on a biometric device for years and never generate an identical template.

Biometric matching is the automated comparison of biometric templates to determine their degree of similarity or correlation. A match attempt results in a score that, in most systems, is compared against a threshold. If the score exceeds the threshold, the result is a match; if the score falls below the threshold, the result is a non-match.

Biometric matching takes place through algorithms that process biometric templates. These algorithms utilize data contained in the template in order to make valid comparisons,

Formatted: Bullets and Numbering

Formatted: Bullets and Numbering

Deleted: Draft 1.0

accounting for variations in submission. Without the vendor algorithm, there is no way to compare biometric templates – comparing the bits which comprise the templates does not indicate if they came from the same user.

The matching process involves the comparison of a match template, created upon sample submission, with the reference template(s) already on file. In 1:1 applications, there is generally a single match template matched against one or more reference templates associated with a given user. In 1:N identification systems, the one or more match templates may be matched against millions of reference templates. Biometric systems do not provide 100% matches, though systems can provide a very high degree of certainty. An identical match is an indicator that some sort of fraud is taking place, such as the resubmission of an intercepted or otherwise compromised template.

A **score** is a value indicating the degree of similarity or correlation of a biometric match. Traditional authentication methods – passwords, PINs, keys, and tokens - are binary, offering only a strict yes/no response. This is not the case with most biometric systems. Nearly all biometric systems are based on matching algorithms that generate a score subsequent to a match attempt. This score represents the degree of correlation between the verification template and the enrollment template. There is no standard scale used for biometric scoring: for some vendors a scale of 1-100 might be used, others might use a scale of –1 to 1; some vendors may use a logarithmic scale and others a linear scale. Regardless of the scale employed, this verification score is compared to the system's threshold to determine how successful a verification attempt has been. Match scores can be associated with a probability that two pieces of biometric data are from the same individual.

A **threshold** is a predefined number, often controlled by a biometric system administrator, which establishes the degree of correlation necessary for a comparison to be deemed a match. If the score resulting from template comparison exceeds the threshold, the templates are a "match" (though the templates themselves are not identical). When a biometric system is set to low security, the threshold for a successful match is lower than when a system is set to high security.

Formatted: Bullets and Numbering

Deleted: Draft 1.0

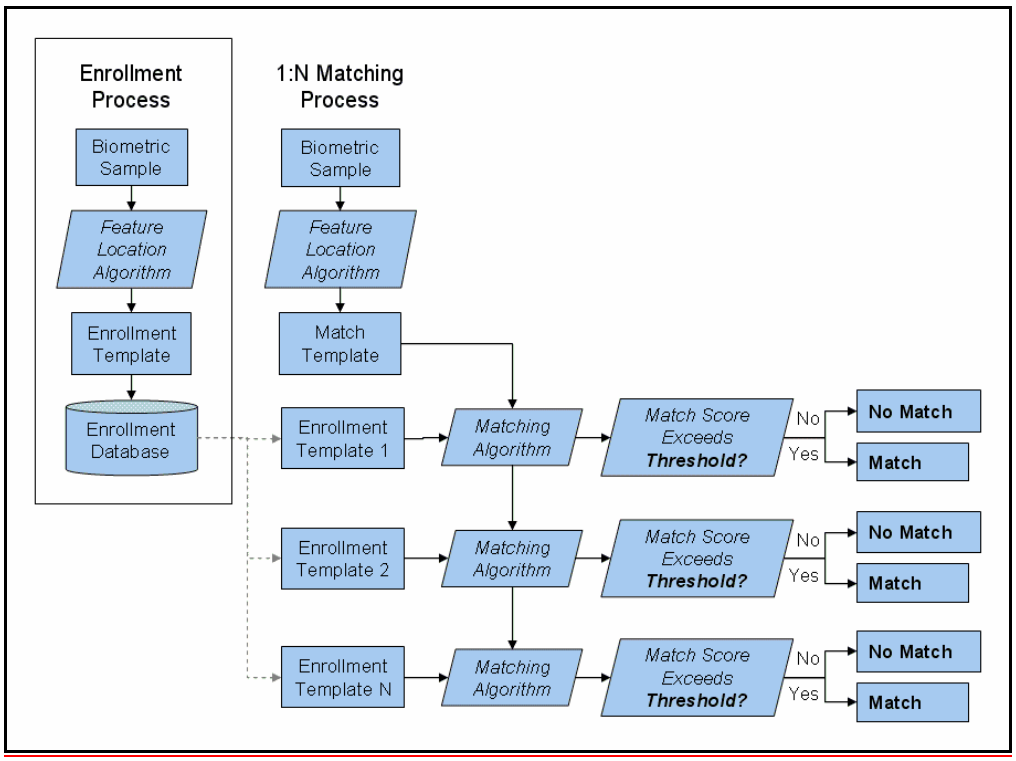


Figure 25: Use of Thresholds in 1:N Matching Process

Formatted: Bullets and Numbering

A **decision** is the result of the comparison between the score and the threshold. The decisions a biometric system can make include match, non-match, and inconclusive, although varying degrees of strong matches and non-matches are possible. Depending on the type of biometric system deployed, a match might grant access to resources, a non-match might limit access to resources, while inconclusive may prompt the user to provide another sample.

An **attempt** is the submission of a biometric sample on the part of an individual for the purposes of enrollment, verification, or identification in a biometric system. An individual may be permitted several attempts to enroll, to verify, or to be identified.

Deleted: Draft 1.0

Appendix C : Biometric Identification Capabilities Questionnaire

This questionnaire was issued to biometric developers and integrators in order to gather their feedback on issues central to the execution of 1:300m matching.

Deleted: is

Deleted: ,

ROM = Rough order of Magnitude

1. Proven Scalability/Accuracy of Technology

These questions are designed to gather information on technology scalability and accuracy as proven through deployments (exclusive of filtering)

1.1. Scale of Largest Operational Deployments
1.1.1. Size of current and target maximum databases
1.2. Accuracy
1.2.1. FMR/FNMR/FTE
1.3. Throughput
1.3.1. Number of transactions per day
1.3.2. Peak transactional capability
1.3.3. Average response time and typical range of responses (shortest to longest)
1.4. Availability
1.4.1. Operating hours/timeframe/limitations
1.5. Relation between Accuracy/Throughput/Availability
1.6. Enrollment Requirements
1.6.1. Average length of time required to enroll
1.6.2. Type of text and biometric data captured
1.6.3. Impact of enrollment loads on scalability
1.7. Application of This System (e.g. Voter Registration, National ID, Etc.)
1.7.1. Population composition

2. Theoretical Scalability/Accuracy of Technology

These questions are designed to gather information on technology scalability and accuracy as projected through tests or theoretical assessments (exclusive of filtering)

2.1. Projected Accuracy for a 1:300m System
2.1.1. FMR/FNMR/FTE
2.1.2. Estimate of accuracy necessary to achieve suitable uniqueness determination performance in DL application
2.2. Projected Throughput
2.2.1. Number of transactions per day
2.2.2. Peak transactional capability
2.2.3. Average response time and typical range of responses (shortest to longest)
2.3. Availability
2.3.1. Operating hours/timeframe/limitations
2.4. Bases for Accuracy, Throughput, Availability Estimates
2.5. Enrollment Functions Required To Achieve Suitable Performance
2.5.1. E.g., for fingerprint systems: number of fingers required; flat vs. rolled acquisition; type and resolution of sensors; degree of supervision and direction required; enrollment time
2.5.2. E.g., for facial systems: type of sensor; lighting requirements; enrollment time
2.5.3. User functions required to achieve suitable performance
2.6. Functions Typically Performed By Operator

Deleted: Draft 1.0

2.6.1. E.g., type/size of gallery of potential matching records presented to operator for decision/resolution
2.6.2. Biometrics-related functions (if any) required to be performed by operator
2.6.3. Training/expertise level of operator required to perform system functions
2.6.4. How much does operator expertise affect accuracy/throughput of system

2.7. Costs

2.7.1. ROM cost for a system capable of conducting 1:300M uniqueness determination
2.7.2. ROM cost for core matching capabilities
2.7.3. ROM cost per workstation; quantity 1000; quantity 10,000

3. Potential Scalability/Accuracy with Filtering

These questions are designed to gather information on the ability of a technology to scale accurately to perform 1:300m identification through use of exogenous filters

3.1. Suitable/practical means for filtering database

3.1.1. Gender, age, geography, etc.
3.1.2. Characteristic(s) of biometric that makes a particular filter effective
3.1.3. Tradeoffs likely to result from employing one or a combination of filters
3.1.4. Field deployments/objective tests that demonstrate practicality/effectiveness of various filters
3.1.5. Estimate of price/performance/accuracy improvements achievable via use of filters

3.2. Costs

3.2.1. ROM cost for a system capable of conducting 1:300M uniqueness determination with filtering

4. Potential for Performance Optimization through Multi-Modal Deployment

These questions are designed to gather information on the ability of a technology to scale accurately through use with other biometric technologies

4.1. Degree to Which Performance May Be Improved Through Use with One or More Additional Biometrics

4.1.1. Which other biometrics are most practical for multi-modal application and why
4.1.2. Impact on performance (accuracy, throughput, etc.) of employing multiple biometrics

4.2. Cost Impact of Employing Multiple Biometrics

4.3. Potential Technical and/Or User Interface Issues Employing Multiple Biometrics

5. General Data on This Technology

These questions are designed to gather information on other key factors related to current and future scalability and deployability

5.1. Ongoing Research into Core/Peripheral Technologies Likely To Significantly Improve Performance/Accuracy within Next 18-36 Months

5.2. Population Sets Known To Present Practical Problems for This Biometric

5.2.1. Recommended mitigation strategies

5.3. Common User/Operator Errors Associated With This Biometric

5.3.1. Recommended mitigation strategies

5.4. Common Spoofing/Uncooperative User Strategies Associated With This Biometric

5.4.1. Recommended mitigation strategies

Deleted: Draft 1.0

Page Break

Biometric Technologies Capable of 1:N Matching

Facial recognition, fingerprint, and iris recognition are the three technologies evaluated for suitability for a 1:300m application.

Not included in this list of 1:N technologies are voice-scan (or speaker verification) and DNA verification. Voice-scan technology can be used for small-scale identification, on the order of dozens or hundreds of records, and has been leveraged for criminal forensics-style applications. However we are currently unaware of any deployments in which the technology has been deployed to databases of tens or hundreds of thousands of enrollees. In addition, the behavioral nature of voice-scan is such that attempts to mask or alter one's voice in a fashion not perceptible by 3rd parties may further undermine the ability to execute 1:N identification. As data emerges from studies or deployments in which voice-scan performs large-scale identification, the technology may warrant further evaluation. DNA is not yet developed as an automated identification technology in the biometric identification sense; manual processing is still necessary. It can also be argued that DNA identification is based less on measurement of a characteristic than on the material retention of a physical sample. However, DNA identification technology may eventually come to resemble biometric identification as currently implemented, such that the technology could be used to perform large-scale automated identification.

The following discussion of facial recognition, fingerprint, and iris recognition technologies' core operations, strengths, weaknesses, and maturity provides a framework for subsequent discussions of tests, deployments, and provider data.

Page Break

Facial Recognition

Facial recognition technology is based on features such as the location and composition

of distinctive features of the face, as well as the spatial interrelation between the eyes, nose, mouth, cheekbones, chin, and forehead.

Strengths for 1:300m Identification
High availability reduces FTE rate (provides near-universal enrollment)
Ability to use multiple matching algorithms to execute 1:N matching; matching algorithms are normally sensor-independent
History of 1:N deployment against large databases
Little to no training required to utilize acquisition devices
Capable of rapid 1:N searches
Compatible with databases of facial images
Weaknesses for 1:300m Identification
Accuracy issues: distance, angles, lighting, time all impact technology's ability to identify individuals
Changes in hairstyle, facial hair reduce matching capabilities
Performance differs by ethnicity
1:N searches generally result in candidate lists which must be manually reviewed

Figure 2: Facial Recognition in Identification Systems

Typical Uses

Facial recognition is deployed in large-scale civil ID applications (such as drivers' licensing and voter registration), surveillance applications, law enforcement applications (such as booking stations), and casinos. It is most often deployed in 1:N applications, searching databases of facial images for close matches and returning lists of likely suspects. Increased use of facial recognition has occurred in large-scale ID projects in which facial imaging already takes place and the technology can leverage existing processes and data.

Strengths and Weaknesses

Facial recognition's significant advantage in a DL/ID environment is its ability to be acquired as part of an existing process (facial photography). The technology's near zero-effort acquisition ensures near total enrollment with almost no impact on current acquisition processes. However, such considerations are out of the scope of the Phase I Report, warranting consideration only if the core technology is shown capable of 1:N identification.

Facial recognition has been shown susceptible to high error rates,

particularly FNMR, when any of several variables are introduced into the matching process. Such variables include the time lapsed between enrollment and matching, the angle of facial acquisition, lighting, distance, facial hair, and glasses.

Facial Recognition Processes

Acquisition

Facial recognition technology can acquire faces from almost any static camera or video system that acquires sufficient quality and resolution images. For optimal performance, images will be acquired through high-resolution cameras, with users directly facing the camera and with low-intensity frontal lighting of the face, as is consistent with 1:N multiple enrollment detection systems. Such systems utilize controlled and consistent enrollment environments: users are required to stand or sit at a fixed distance from a camera, with fixed lighting and a fixed background.

Image Processing

Most facial recognition image processing entails cropping and normalizing images to a consistent size and orientation, converting color images to black and white, and in many cases location of landmarks to assist in automated matching. Characteristics such as the middle of the eyes are used as points of reference. Since most facial recognition systems acquire multiple face images to enroll individuals – from as few as three images to well over 100, depending on the vendor and the matching method – rapid image processing routines are essential to system operations.

Distinctive Characteristics

The features most often utilized in facial recognition systems are those least likely to change significantly over time: upper ridges of the eye sockets, areas around the cheekbones, sides of the mouth, nose shape, and the position of major features relative to each other. Areas susceptible to change or obscuration - such as areas of the face immediately adjacent to a hairline - are usually not relied upon for identification. One of the challenges involved in facial recognition technology is that the face is a reasonably variable physiological characteristic. As opposed to a fingerprint, which might be scarred but is difficult to alter dramatically, or the iris, which is reputedly stable for one’s entire life, faces can be changed voluntarily enough to reduce a system’s matching accuracy.

A user who smiles during enrollment and grimaces during verification or identification is more likely to be rejected than one who does not intentionally alter his or her expression during authentication. Behavioral changes such as alteration of hairstyle, changes in makeup, growing or shaving facial hair, adding or removing eyeglasses can impact the ability of facial recognition systems to locate distinctive features.

Template Generation and Matching

Enrollment templates are normally created from multiple processed facial images, although a single image can be used. Multiple facial images are preferred for generation of enrollment templates, as this provides an increased opportunity to “match” the facial positioning of enrollment with that of identification. Templates can vary in size from less than 100 bytes for rapid searching to over three kilobytes. Instead of a single record being returned, in most cases a candidate list with a number of potential matches is returned in large-scale facial recognition identification. For example, a system may be configured to return the ten most likely matches on a search of a 10,000-person database. A human operator would then determine which if any of the ten potential matches were actual matches. The degree to which system performance varies when using a single facial image (as is consistent with DL/ID concepts of operation), as opposed to several facial images acquired with slightly different facial aspects, has not to our knowledge been studied.

Competing Facial Recognition Technologies

A handful of facial recognition technologies compete within the biometric market. Solutions may be based on global features, leveraging the common eigenfaces method of rendering facial images by combining features from a database of facial images. Other solutions are based more directly on localized features. Other solutions are based on spatial ratios and relationships between certain fixed points on the face.

Efforts are currently underway within U.S. and international standards groups (ICAO TAG NTWG, ISO/ IET JTC1 SC 37, INCITS M1 Biometrics) to standardize the manner in which facial images are acquired and utilized for biometric matching. When using standard image capture and landmark location processes (e.g. positioning the eyes at a fixed distance for all users), images are more likely to be usable by multiple matching algorithms.

Fingerprint

Fingerprint biometrics is based on the ridges, valleys, ridge endings, loops, whorls, and other distinctive features found on the human fingerprint. Biometric systems based on fingerprints can be divided into fingerprint, used in 1:1 systems, and AFIS, used in 1:N systems (the differences between these types of systems are discussed below).

Strengths for 1:300m Identification
Based on distinctive characteristics
Availability of multiple samples (ten fingerprints) increases overall accuracy and scalability
Mature and widely accepted standards in DPI, resolution provide framework for consistent technology usage
Strong competition in market drives emergence of new solutions
Weaknesses for 1:300m Identification
Fingerprint quality varies by age, race; subject to wear and tear (both incidental and intentional)
Percentage of users unable to enroll due to fingerprint quality
Accuracy can diminish over time
Sensor surfaces can be scratched, require maintenance
Scalability of multiple-finger civil AFIS (flat fingerprint) systems uncertain
Highly contingent on quality of initial data

Figure 3: Fingerprint in Identification Systems

Typical Uses

Fingerprint technology is used by hundreds of thousands of people daily to verify availability for public services, to access networks and PCs, to enter restricted areas, and to authorize transactions. The technology is used broadly in a range of vertical markets and within a range of horizontal applications, primarily PC / Network Access, Physical Access / Time and Attendance, Civil ID, and Criminal ID.

Including OEMs and application developers, there are over 100 companies operating in the fingerprint marketplace. Approximately half of these companies are “core technology” firms, meaning that they manufacture or develop a basic component of the fingerprint system such as a sensor or an algorithm. These companies may utilize a proprietary sensor technology or use sensors from another manufacturer. In addition, these core technology companies may develop or manufacture the following components:

Sensors and modules (designed for integration into 3rd party devices or peripherals)
Devices and peripherals (ready-to-deploy units for logical and physical access)
Algorithms (perform extraction and matching functions)
Application software (enable PC or network access)

Fingerprint technology can also be divided according to minutia-based and pattern-matching vendors. Approximately 75% of fingerprint solutions use minutia-based extraction algorithms, which generate and compare templates based on the relative position and direction of dozens of ridge endings and bifurcations found in fingerprints. Pattern matching algorithms are based on the regional characteristic present across multiple ridges as opposed to single points.

AFIS technology utilizes groups of fingerprints, as well as fingerprint classification methods, to enable large-scale searches against databases of thousands to millions of fingerprints. AFIS was first developed in the mid to late 1970's as a means of automating manual searches required to identify a fingerprint from the tens of millions of fingerprints in FBI files. Today, the term "AFIS" refers to the *technology* used to automate large-scale fingerprint searches as well as the *industry* inclusive of automated acquisition, storage, and identification of fingerprints.

From a technology perspective, the AFIS industry can be viewed as two distinct segments: (1) fingerprint acquisition and (2) fingerprint storage and processing. Fingerprint acquisition components collect high-quality electronic fingerprint images, either directly from placement of fingerprints on "live-scan" readers or from scanned ink cards. Fingerprint storage and processing components generate fingerprint templates and perform 1:N matching against fingerprint databases. Mature standards define the capture resolution, permissible distortion, and compression ratios of fingerprint images, ensuring interoperability of images across AFIS vendor systems.

Fingerprint Processes

Acquisition

Image acquisition is a major challenge for fingerprint providers, as fingerprint quality varies substantially from person to person and from fingerprint to fingerprint. Certain populations are more likely than others to possess faint or difficult-to-acquire fingerprints, whether due to occupational wear and tear or physiology. In addition, environmental factors can impact image acquisition: in cold weather, fingerprint oils (which improve

image quality) dry up, such that fingerprint images may appear faint. Sensor size can also impact a system's accuracy and performance. Very small sensors acquire a smaller portion of the fingerprint, such that less data is available to enroll and match templates. Users with large fingers may find it difficult to place their fingerprint in a consistent fashion, leading to false non-matches.

Image Processing

Image processing subroutines convert the fingerprint image's gray pixels to white and black. A series of thick black ridges (the raised part of the fingerprint) results, contrasted to white valleys. The ridges are then "thinned" down to a single pixel in width to enable precise location of features.

Location of Distinctive Characteristics

The fingerprint is comprised of ridges and valleys which form distinctive patterns, such as swirls, loops, and arches. Many fingerprints also have a core, a central point around which these patterns curve. Points found at the lower left or right corner of the fingerprint around which ridges are centered in a triangular shape are known as deltas. Discontinuities and irregularities in ridges and valleys – known as minutiae – are the features upon which most fingerprint technologies are based. The primary minutia types are ridge endings (the point at which a ridge ends) and bifurcations (the point at which a ridge splits). Depending on the size of the sensor and the sensitivity of the algorithm, a typical fingerprint may produce between 30 and 50 minutia – larger platens acquire more of the fingerprint image such that more minutiae can be located. This information is fairly stable throughout one's life, and differs from fingerprint to fingerprint.

Fingerprints contain sufficient information to enable large-scale identification using multiple fingerprints. However, conducting such large-scale searches is time consuming and processor-intensive. To limit the number of fingerprints that must be searched in identification systems (and thereby to limit search time and processing demands), AFIS technology also classifies the group of fingerprints acquired from each individual according to prints' global characteristics. An AFIS may therefore only need to search that percentage of records whose group classification matches that of the enrollee, instead of searching an entire fingerprint database. The percentage of an AFIS database that must be searched subsequent to classification is referred to as a "penetration rate" – lower penetration rates result in faster, more accurate searches. Collecting more

fingerprints per individual results in a lower penetration rate. However, it is critical that fingerprints be classified correctly, or else an imposter may go undetected in a 1:N search. Newer AFIS systems utilize matching methods that do not rely on this type of classification scheme.

Page Break

Template Generation and Matching

Vendors utilize proprietary algorithms to locate fingerprint minutiae. Information used when mapping minutiae can include the location and angle of a minutia point, the type and quality of minutia, and the distance and position of minutiae relative to the core. Fingerprint images may contain distortions and “false minutiae” that must be filtered out before template creation; scars, sweat, and dirt can appear as minutiae. Algorithms scan images and eliminate anomalous features that seem to be in the wrong place, such as adjacent minutiae or a ridge crossing perpendicular to a series of other ridges. A large percentage of false minutiae are discarded in this process, ensuring that the template generated for enrollment or matching accurately reflects legitimate biometric data.

Fingerprint templates can range in size from approximately 200 bytes to over 1000 bytes (in contrast to the 10-20 kilobytes required to store a single compressed fingerprint image). These templates cannot be “read” like a fingerprint image. Matching algorithms are required to process templates and to determine the correlation between the two.

Sensor types

Fingerprints are acquired through optical, silicon, and ultrasonic sensors.

Optical technology is the oldest fingerprint imaging technology and the most widely used for 1:N systems. Optical technology has several strengths: proven reliability over time, resistance to electrostatic discharge, resolutions of 500 dpi and above, and large surface area (certain optical sensors can acquire multiple fingerprints in a single placement). Weaknesses include size and power constraints – the platen requires more surface area and depth than silicon technology, such that optical technology cannot e built into very small devices – and difficulty acquiring dry fingerprint images.

Silicon technology uses coated chips to image fingerprints. Most silicon fingerprint technology is based on capacitance, wherein the silicon sensor acts as one plate of a capacitor and the finger acts as the second plate. Silicon sensor strengths include image

quality (approaching that of many optical devices), modest size and power requirements (such that sensors can be integrated into small, low-power devices), and low cost. Silicon sensor weaknesses include durability, susceptibility to electrostatic damage, and performance in challenging conditions. Silicon sensors are primarily deployed as logical access solutions, have found limited but increasing deployment in physical access deployments, and are rarely deployed in high-traffic public sector applications such as DL/ID.

Ultrasonic devices generate fingerprint images by emitting ultrasonic waves, measuring the “echo” returned when the acoustic waves meet the ridges and valleys of the fingerprint. Among the advantages of this method are that ultrasonic devices are better able to acquire images from low-quality fingerprints with surface contaminants – dirt, grease – that normally reduce image quality. Theoretically, better image quality will result in more accurate fingerprint matching, though there have been no tests that demonstrate ultrasonic scanner’s compatibility with multiple vendor’s matching algorithms. The platen used in ultrasonic technology does not require a special coating, as is often the case in optical imaging. Disadvantages of the ultrasonic imaging method include a bulky device size, the presence of moving parts necessary to image the fingerprint, as well as the length of time required to gather images during enrollment. Ultrasonic technology can be used in civil ID applications such as DL/ID.

Iris Recognition

Iris recognition technology is based on the ridges, furrows, and striations that characterize irises.

Strengths for 1:300m Identification
Extremely low False Match Rates Thought to be a highly stable and distinctive physiological characteristic, unchanging over time Ability to leverage both irises, which have been shown in vendor tests to be highly independent
Weaknesses for 1:300m Identification
Scalability in 1:N applications not yet proven in field evaluations Enrollment can be problematic for certain users Acquisition of iris can be time consuming and difficult

Figure 4: Iris Recognition in Identification Systems

Overview

The iris is a highly distinctive characteristic, reportedly stable at birth and unchanging through one's life, and has not been shown to be alterable in any fashion (although the appearance of an iris can possibly be masked through use of certain contact lenses). Iris recognition technology is primarily deployed in high-security physical access implementations, but has found increasing acceptance in travel and transportation and civil ID applications. One company (Iridian) holds the key patents for utilizing the iris for identification. Iridian both licenses its core technology to integrators and developers and it markets systems directly to deployers. Although alternative implementations of iris recognition technology have recently begun to emerge, Iridian has historically been very aggressive about defending its patent rights.

Iris Recognition Processes

Image Acquisition

Iris recognition matching requires the acquisition of a high-resolution image of the eye, illuminated by an infrared imager, in order to effectively map the details of the iris. The acquisition process, and the effort required on the part of the user, differs according to the type of acquisition device used. Primary iris recognition imaging systems include kiosk-based systems, physical access devices using motorized cameras, and inexpensive desktop cameras. Although iris recognition vendors do not emphasize their use of infrared light, each system does rely on infrared imaging using wavelengths in the 700-900nm range (judged to be safe by the American Academy of Ophthalmology).

Depending on the quality and positioning of the acquisition device, and the level of training and supervision granted to the enrollee, acquisition of iris images requires moderate to high levels of training and attentiveness. Users must be cognizant of the manner in which they interact with the system, as enrollment and matching require fairly precise positioning of the head and eyes. Also, users with poor eyesight, or those incapable of lining up their eye with the technology's guidance components, have difficulty using the technology.

Image Processing

After the eye is located, algorithms locate the iris' outer and inner borders. Locating the iris-pupil border can be challenging for users with very dark eyes, as there may be very

little difference in shade as rendered through 8-bit grayscale imaging. Once the parameters of the iris have been defined, a black and white image of the iris is used for feature extraction. The core technology can account for pupil dilation, eyelid occlusion, and reflections due to the acquisition camera. When the pupil dilates, the iris patterns shrink and expand in a normalized fashion such that algorithms can translate a dilated match to a non-dilated enrollment.

Page Break

Distinctive Features

Iris recognition technology uses a horizontal band extending from the far left to the far right of the iris for feature extraction. The patterns that comprise the visual component of the iris are highly distinctive. The trabecular meshwork, a tissue that gives the appearance of dividing the iris in a radial fashion, is the primary distinguishing characteristic. Other visible characteristics include rings, furrows, freckles, and the corona. Tests have shown that individuals' left and right eyes have different iris patterns, and that even identical twins' irises have almost no statistical similarity. Iris recognition algorithms map segments of the iris into hundreds of independent vectors. The characteristics derived from iris features are the orientation and spatial frequency of distinctive areas along with the position of these areas.

Template Generation and Matching

The vectors located by the iris recognition algorithm are used to form enrollment and match templates, which are generated in hexadecimal format as opposed to binary. Depending on the iris recognition solution, between one and four iris images may need to be captured for enrollment template generation. The use of multiple images ensures that the data extracted to form a template is consistent. Iris recognition solutions generally perform identification as opposed to verification, meaning that the match template is compared against multiple enrollments until a match is located. The matching process is very rapid, with hundreds of thousands of records searched per second.

Multimodal Solutions

The challenges involved in scaling biometric systems to execute 1:300m identification in a DL/ID environment may be mitigated through the use of multimodal biometric systems. By leveraging combinations of technologies such as facial recognition, fingerprint, and

iris recognition, multimodal systems may be capable of providing a higher degree of accurate scalability than any single biometric technology. Such an approach may be the only solution to the 1:300m challenge. In order to determine if and how multimodal systems can be used to improve accuracy and enrollment capabilities for the purposes of 1:300m identification, it is necessary to understand the basic operations and categorizations of multimodal systems.

Definition

Multimodal biometric systems are those which utilize, or collect for the purpose of utilizing, more than one physiological or behavioral characteristic for enrollment, verification, and/or identification. Multimodal systems address specific problems found in monomodal biometric systems, including the following:

Biometric systems are subject to false match and false non-match errors. Depending on the type of biometric system deployed, excessive matching errors can lead to security breaches, undetected fraud, and processing delays.

Biometric systems are subject to failure to enroll and failure to acquire errors. Such errors can be attributable to lack of required physiological characteristics, to insufficiently distinctive biometric characteristics, or to an inability to adhere to device interaction requirements. FTE and FTA errors result in a percentage of individuals permanently or temporarily unable to use a given biometric system. This creates problems for deployers, as a backup authentication method must be maintained, and malicious users may intentionally fail to enroll in order to attack the weaker authentication method (e.g. password).

Recent demonstrations have shown that many biometric systems can be fooled by non-live data, some with little effort, others with substantial effort. This raises the possibility that difficult-to-repudiate transactions could be created and associated with an individual without his awareness, and that individuals could easily circumvent 1:N detection through use of fraudulent data.

\Multimodal biometric systems are designed to provide the following benefits:

Reducing false non-match rates and false match rates. By deploying more than one biometric technology for 1:N and/or 1:1 processing, and intelligently combining or fusing match results from both systems, it is possible to reduce the overall system’s matching error rates. For example, to reduce 1:1 false matching, a multimodal

system can provide two subsystems against which a user must match in order to defeat the system. Similarly, to reduce 1:1 false non-matching, a multimodal system may only require that a user match against one of two subsystems.

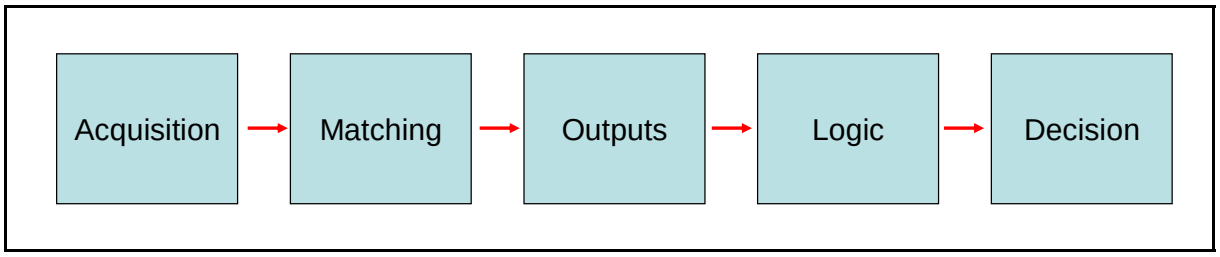
Providing a secondary means of enrollment, verification, and identification for users unable to enroll and/or authenticate through a primary biometric technology. By deploying more than one biometric technology, a deployer can ensure that a higher percentage of individuals are enrolled and matched in a biometric system, reducing the need for fallback or secondary processing. This in turn can reduce costs and security risks.

Combating attempts to spoof biometric systems through non-live data sources such as fake fingers. By implementing a multimodal system that requires dual authentication, an attacker must successfully spoof both systems, or spoof one while attempting to match as an imposter in another.

These direct benefits can lead to indirect benefits such as increased system security, reduced deployment costs, and increased ease of use. However, realizing benefits from a multimodal biometric system requires intelligent combination and utilization of technologies, devices, and algorithms; requires an application for which a multimodal solution is both viable and useful; and requires effective design of enrollment and matching processes. Issues and variables involved in designing multimodal biometric processes, selecting multimodal technologies, and identifying suitable deployment environments for multimodal solutions are addressed in the following sections.

Concepts of Operations

Multimodal biometric systems can take a variety of forms with wide-ranging levels of complexity. In order to provide a method of categorizing the major functional elements of multimodal systems, and to allow for a common baseline for discussions of multimodal concepts of operation, a five-stage framework, incorporating (acquisition, matching, output, logic, and decision, can be used. Variables within multimodal biometric systems can be categorized according to the major steps in the biometric processing cycle:



Acquisition Process variables apply to collection and assessment of biometric data in a 1:300m system.

Matching Process variables apply to comparison of biometric data in a 1:300m system.

Output Process variables apply to generation of match results through biometric comparisons in a 1:300m system.

Logic Process variables apply to methods of utilizing one or more sets of match results in a 1:300m system.

Decision Process variables apply to decisions that result from biometric transactions or search events in a 1:300m system.

This approach allows us to (1) isolate areas that inform multimodal performance and (2) allow for inclusion of various modes of multimodal system implementation, from simple to complex.

Multimodal Acquisition

A wide range of variables apply to the collection of biometric data during enrollment and/or identification. This collection event, and the parameters around the event, falls under the auspices of *Acquisition* variables. In nearly all multimodal biometric systems more than one acquisition device will be necessary, as each biometric discipline utilizes its own acquisition technology: cameras, scanners, microphones, etc. Although a multimodal system may be designed such that a primary biometric is used for most transactions and a secondary biometric is used for exception cases, in a DL/ID environment, it is assumed that both systems will be utilized. This directly impacts the manner of acquisition of biometric data. *Acquisition* variables are particularly important due to their direct impact on enrollees and end users. These variables may impact transaction times, storage requirements for both capacity and flexibility, and level of supervision required for system operation. *Acquisition* variables are also important due to their relation to limiting FTE and FTE rates. With sufficient time and effort, an enrollment operator can drive FTE down to near-zero for many technologies; however,

this does not address the needs of a transactional system in which a user may need to provide data in a time-constrained, unattended usage environment. In this environment, users may enroll successfully but be incapable of authenticating on subsequent interactions with the system. The following table describes the primary *Acquisition* variables and lists ranges associated with these variables.

Acquisition Variable	Description	Range
Number of Acquisition Devices (E, M)	**Number of acquisition devices involved [maximum].** The number of acquisition devices in a multimodal system can impact system costs and deployment footprint.	1-N devices utilized
Number of Core Technologies (E, M)	**Number of core technologies involved [maximum].** In most cases, each core technology (e.g. fingerprint, facial recognition) will be associated with a single acquisition device. However in some cases a single device may be capable of acquiring more than one biometric type, e.g. iris and face.	1-N technologies utilized

Figure 5: Acquisition Variables

Multimodal Matching

Nearly all multimodal systems incorporate multimodal *Matching Processes*. The primary exception is a multimodal system in which more than one biometric is acquired on enrollment (e.g. as future placeholders for emerging systems) but in which only one biometric is used on an ongoing basis for 1:N and 1:1 functionality. Note that matching is separate from output and logic, being applicable solely to matching processes. In almost all cases matching processes are black-box, meaning that the actual protocols and mechanics associated with template generation and comparison are not known beyond the vendor. The following table describes the primary *Matching* variables and lists ranges associated with these variables.

Matching Variable	Description	Range
Number Of Core Technologies	**The number of core technologies involved.** Multimodal matching processes are directly impacted by the number of core technologies involved. Because biometric matching is, for many technologies, a processor-intensive function, multimodal systems may be designed to only match secondary biometrics in certain situations. A system in which three different biometrics are available may only ever match two.	1-N technologies utilized
Number Of Algorithms	**The number of matching algorithms involved.** Multimodal systems must, with rare exceptions, utilize a different matching algorithm for each biometric characteristic acquired (e.g. face, finger). At the same time, multimodal systems can be designed to utilize a single matching algorithm for each biometric characteristic, or may utilize multiple algorithms to process a given piece of biometric data. This multi-	1-N matching algorithms utilized

	algorithmic approach in itself contains many potential sub-variables, including weighting of match scores at the logic processes phase. Most systems do not apply a multi-algorithmic to a single technology.	
Source Fusion	**A process in which biometric samples (e.g. identifiable images, voice patterns) are combined in either raw or processed form to create a new meta-sample; this meta-sample may be processed through a different algorithm than either source sample.** Source fusion is a highly emergent process by which more than one biometric type is fused at the sample stage in order to generate a "new" biometric. In most cases a pattern-matching algorithm would be used to process this new type of biometric data.	Data fused from characteristics 1, 2,…N

Figure 6: Matching Variables

Multimodal Output

Outputs that result from matches in multimodal systems are identical to those in monomodal systems. For 1:300m identification systems outputs include Rank 1 and Rank N results. Match responses from each component of a multimodal system are critical to developing large-scale ID systems such as in a DL/ID environment. The following table describes the primary *Output* variables, and lists ranges associated with these variables.

Output Variable	Description	Range
Rank 1	**Match output rendered as a single candidate.** Open and closed set multimodal systems are each capable of generating *Rank 1* Match outputs, returning a Candidate ID corresponding to the record with the strongest correlation score.	Candidate ID N
Rank N	**Match output rendered as a list of candidates.** Open and closed set multimodal systems are each capable of generating candidate lists indicating the top N matches, or *Rank N matches*, each of which may be associated with a given score or simply ranked in numerical order.	Candidate ID 1, 2, 3…N

Figure 7: Output Variables

Multimodal Logic

Multimodal *Logic* is the focus of considerable active multimodal research, as academics and vendors investigate the most effective method of combining results from multiple systems to improve overall biometric system performance. Determining the most effective way of combining match results from disparate systems is at the core of multimodal systems operation. To date, fielded products have often used rudimentary *and/or* logic in multimodal systems, driven partly by a desire to simplify deployment but also by a lack of strong familiarity with what components' match scores truly indicate.

More advanced approaches have been developed by vendors with access to the core IP behind more than one biometric modality and by integrators with experience in developing solutions that fuse inputs through algorithm layers to generate improved cross-system results.

Several of the following *Logic* areas can be categorized as fusion processes. Fusion systems use specially developed decision models to intelligently combine the results from more than one biometric system into a "master" decision. Fusion biometrics leverage the probabilities associated with biometric match events in driving match/no-match decisions. For example, a strong match on an iris recognition system may require only a very weak match on a parallel-operating facial recognition system, while a weak match on iris recognition may require a much stronger facial recognition match. In order for fusion biometrics to provide demonstrable improvements in accuracy, a developer requires access to detailed information regarding vendor-specific biometric scores: specifically, how scores correspond to probabilities when determining matches and non-matches. The following table describes the primary *Logic Process* variables and lists ranges associated with these variables.

Logic Variable	Description	Range
System Weight	**Multimodal logic in which one system's output is weighted more heavily than another system's.** *System Weight Logic* is generally applicable in systems that combine a strong biometric with a weak biometric. In order to ensure that the stronger biometric technology's decision more directly impacts match decisions, a 75/25 or 80/20 weight may be accorded to a given system, such that an extremely low match score. This logic can only be applied to multimodal systems whose outputs are non-binary (which would include native and standardized match scores as well as probability outputs). This logic is designed to reduce FNMR and FMR.	Weight A (System 1), Weight B (System 2), …Weight N (System N)
Score Weight	**Multimodal logic in which relatively strong match scores figure more heavily in a decision process than low match scores.** *Score Weight Logic* is applicable to multimodal systems that utilize core technologies of roughly equivalent strength, such as fingerprint and iris recognition or facial recognition and voice verification. This logic can only be applied to multimodal systems whose outputs are non-binary (which would include native and standardized match scores as well as probability outputs). This logic is designed to reduce FNMR, built on the assumption that a single strong match should bear a stronger weight in the overall logic.	Weight A (System Score 1), Weight B (System Score 2), … Weight N
Combined Score 'And'	**Logic process in which the application requires a total score across all systems.** This logic process allows for a variety of combinations of match scores, often in conjunction with system weight and/or score weight logics, as above. This logic process allows for more detailed means of incorporating results from two or more systems.	Combined Minimum Score N on Systems A, B

Logic Variable	Description	Range
Match Score 'And'	**Logic process in which the application require a minimum score on all systems.** A more advanced version of Binary Match 'And', this logic process entails that specific match scores known to be associated with desired probability levels be surpassed for each component of a multimodal system. Knowledge of vendor thresholds and how they map to user requirements is a prerequisite of this logic process, designed to reduce FMR.	System A Minimum Score N *and* System B Minimum Score N
Rank N Match 'And'	**Logic process in which the system requires a Rank N Candidate on all systems.** *Rank N Match 'And' Logic* establishes a minimum rank for 1:N searches across all systems within which results for a given individual are assumed to be valid. For example, a Rank 5 Match 'Or' system would return all matches in which an individual was within the top 5 for all systems.	Rank N on Systems 1 and 2 and …N
Rank 1 Match 'And'	**Logic process in which the system requires a Rank 1 Candidate on all systems.** *Rank 1 Rank 1 Match 'And' Logic* requires that a Rank 1 match be attained on all systems for an individual to be declared a match; this process is designed to reduce system FMR.	Rank 1 on Systems 1 and 2 and …N

Figure 8: Logic Variables

Multimodal Decisions

Multimodal *Decisions*, which represent the institution's policies for matching, non-matching, identification, and other functions, are similar to monomodal decision processes with the addition of more complex retry sequences at certain points. Contingent multimodal systems can be designed with 'gates' that implement decision processes subsequent to matching within one biometric technology, with results potentially triggering match decisions or triggering activation of a secondary biometric technology. *Decisions* are closely tied to *Logic*, and represent the institution's actions based on a given match event's ability to successfully address an application's logic requirements. The following table describes the primary *Decision* variables and lists ranges associated with these variables.

Decision Variable	Description	Range
No Match, Terminate Sequence	**Decision process in which no match has occurred, the biometric sequence terminates.** This decision process logically follows a match sequence in which none of the preconditions for matching established at the logic phase are met. If the user has failed to match for N number of transactions, the entire matching sequence may be terminated and revert to manual authentication or user lockout.	No Match
No Match, Retry	**Decision process in which no match has occurred, the biometric sequence is reinitiated.** This decision process is utilized when an individual is granted multiple sequences to authenticate and fails to authenticate within one of the initial sequences.	No Match: Retry N
Match, Grant	**Decision process in which a match has occurred.** This	Match

Access	decision process is a common result in which a user successfully meets the conditions established at the logic phase.	
Rank 1 Match Found	**Decision process in which a candidate has been identified as Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which a Rank 1 match has been found to meet the requirements of the 1:N logic. In most cases the matching record would be returned for further action (investigation, grant of access).	Match, ID#
Rank 1 Match Not Found	**Decision process in which no candidate has been identified as Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which no Rank 1 match has been found that meets the requirements of the 1:N logic.	No Match
Candidate Found Within Ranks 1-N	**Decision process in which a candidate has been identified as being within N positions of Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which an individual is located within N places of Rank 1 in a fashion that meets 1:N logic requirements. In most cases the matching record would be returned for investigation.	Match (rank), ID#
Candidate Not Found Within Ranks 1-N	**Decision process in which no candidate is identified as being within N positions of Rank 1.** This decision process, applicable to ID and watchlist systems, results from a search in which no individual is located within N places of Rank 1 in accordance with 1:N logic requirements.	No Match
Inconclusive	**Decision in which the match result is inconclusive to execute a decision.** This decision process may result in a declaration of no match or in a triggering or a retry cycle.	No Match / Retry N

Figure 9: Decision Process Variables

Multimodal Technology Combinations

Identification (1:N) Technology Ratings and Combination Assessments

The following ratings assess 1:N biometric technologies relative to above criteria.

5: Strongest capabilities
4: Strong capabilities
3: Moderate capabilities
2: Weak capabilities
1: Weakest capabilities

	Accuracy	Ease of Acquisition	Availability	Measurability	Legacy DB
Fingerprint	**5**	**4**	**4**	**5**	**4**
Facial Recognition	2	5	5	5	5

Figure 10: Identification Technology Combinations

Based on these ratings, as well as acquisition synergy considerations, the most promising multimodal combination for 1:N applications will be Fingerprint and Facial Recognition, followed by Iris Recognition and Facial Recognition and then Fingerprint

and Iris Recognition. Limitations and advantages of these 1:N multimodal technology combinations are as follows.

Fingerprint and Facial Recognition

The use of fingerprint and facial recognition as a 1:N multimodal solution provides advantages in terms of ability to leverage legacy databases and (in the case of facial recognition) to leverage existing processes such as photo capture. Because facial recognition has been sanctioned by ICAO as the primary interoperable biometric technology for passport usage, it is likely that substantial effort will be dedicated to developing multimodal solutions that utilize this mandatory data piece in addition to more reliable identifiers such as fingerprint and iris recognition. In terms of 1:N functionality, facial recognition's ability to function as a gross classifier allows it to reduce the size of large 1:N databases and to effect more rapid 1:N fingerprint searches. Executing parallel full-scale 1:N searches through both fingerprint and facial recognition is unlikely to prove highly beneficial, as the results from the facial recognition will not be reliable, even on small databases.

This underscores a challenge of this technology combination: if fingerprints cannot be obtained from a given subject, then only facial recognition can be utilized, meaning that a much less robust 1:N search would need to be executed. Furthermore, a deployer cannot simply assume that a search is required only on a database of individuals unable to enroll in the fingerprint system. A motivated individual can mar his or her fingerprints such that he or she could fail to enroll after having enrolled previously. Therefore facial recognition must be optimized to work as a standalone system to avoid providing imposters with simple workarounds.

Iris Recognition and Facial Recognition

Iris recognition and facial recognition bring the substantial advantage of being capable of being imaged through a single user process and through a single-housing acquisition device (which may contain two separate imaging elements). Therefore many of the process-driven impediments that face multimodal systems are not present in this technology combination. Narita Airport in Tokyo is testing this exact technology combination to determine the efficacy of the combined technologies. The limitation of this technology combination has to do with the relatively marginal role that facial recognition can play in a system with a strong and highly available biometric such as iris

recognition. In most cases, if the iris can be reliably imaged, then the facial recognition components will add very little, such that the cost/benefit of collecting and maintaining such data can be called into question.

Also, since iris recognition is always deployed as a day-forward solution as opposed to a solution that leveraged legacy data, there is less need to address existing large-scale databases. In extremely large-scale 1:N systems, facial recognition could serve as a gross classifier to reduce the demands on the 1:N iris search, as iris technology has not been deployed in highly scaled application environments (those with 1m+ enrollees).

Fingerprint and Iris Recognition

For systems in which certainty regarding match results is an absolute necessity, fingerprint and iris recognition offer similar capabilities in terms of reliable 1:N matching. Fingerprint is more proven in real-world applications, and has been shown to scale with large loads of applicants; iris recognition is harder to circumvent than fingerprint technology, is more universally available, and provides high levels of accuracy. Each technology offers multiple samples, increasing scalability and accuracy. It is likely that for many deployers, iris recognition and fingerprint represent similar-enough capabilities that using both will be excessive. Neither is designed to provide the rapid, inexpensive database-reducing 1:N gross search functions of facial recognition.

Page Break

Test Efforts

Page 68: [2] Deleted **Michael Thieme** **8/1/2003 6:41:00 AM**

Iris Recognition Testing

Of the three technologies under consideration, less independent testing has been conducted on iris recognition than any other. One of the major impediments to iris recognition testing in comparison to facial recognition and fingerprint is the lack of legacy databases against which to execute matching. However, the increased interest in facial recognition technology on the part of government agents for applications such as border control is likely to result in a vast increase in independent testing.

The perception of iris recognition as a highly accurate technology is based on results from internal tests, some of which have taken place on small controlled databases and others of which have taken place on larger operational databases. Iridian, the primary developer and patent-holder of iris recognition technology, has executed a number of internal tests of its technology in order to determine resistance to false matching. The largest and most recent test, entitled Iridian Cross Comparison Test, was published in December 2002.

The following data must be read with the understanding that this testing was executed by the vendor as opposed to an independent body. Independent published iris recognition tests include (1) an operational test executed against a very small (sub-1000) databases by the U.S. Department of Defense

Page 68: [3] Deleted **Michael Thieme** **8/1/2003 6:41:00 AM**

[1] represents a large technology test of the susceptibility of iris recognition to false matching when conducting 1:N matching. This test provides useful information for the evaluation of iris recognition technology for 1:300m identification. The results from this test, by far the largest published iris recognition test (conducted with operational data taken from approximately 120,000 subjects), indicate that iris recognition may be capable of performing identification against very large databases with a very low false match rate. The effective false match rate, according to vendor data, may be less than 1 in 2.79 million for a single enrollee against a database on the order of magnitude of 100 million records. Such performance is well within the bounds established for FMR in discussions with UID9. This is the only test, to our knowledge, that attempts to project performance to a level commensurate with the

[1] *Iridian Cross Comparison Test,* December 2002, available for download at www.iridiantech.com

300m person database under consideration.

However, this test leaves a number of areas unaddressed, and as such can only be judged "inconclusive" with regard to performance against a 300m person database.

Test Population

In Iridian's testing

Page 68: [4] Deleted **Michael Thieme** **8/1/2003 6:41:00 AM**

[2], approximately 120,000 iris templates were collected from various operational databases from around the world. Several gallery databases were created from the following operational iris recognition databases (database sizes approximate):

25,000 enrollees were from U.S. databases

20,000 enrollees were from Middle Eastern databases

65,000 enrollees were from South Asian databases

700 enrollees were from European databases

Iridian then created a 9000 person "probe" database comprised of iris data acquired from Icelandic deployments with which to search the target databases above.

Test Execution

No duplicates enrollments were shared between the probe and the target databases, meaning that any matches that occurred could only be false matches. Large gallery databases were divided into databases of not more than 17,000 enrollees. The probe and target

[2] "Iridian Cross-Comparison Test", December 2002

templates were matched as follows:

9000 x 7374 = 66,366,000 matches

9000 x 16290= 146,610,000 matches

9000 x 17000 = 153,000,000 matches

9000 x 17000 = 153,000,000 matches (etc.)

This totaled 983m matches, a match load roughly equivalent to 1000 new enrollees being matched against a database with 1m enrollees. In the 983m matches executed, the following number of false match errors occurred at various thresholds.

Cross Comparison Testing generated only one metric: the number of false matches at a given threshold (or hamming distance). The Hamming Distance, or HD, represents the degree of similarity between two iris templates. As the HD between two templates grows closer to zero, more similarity is present between two templates. Templates with no similarity would have a HD of 0.50.

Results

Te results of the Cross Comparison Test were as follows.

157 errors (false matches) occurred at the 0.31 threshold (i.e. when the hamming distance between the two matched templates was 0.31 or lower)

32 errors (false matches) occurred at the 0.30 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.30 or lower)

10 errors (false matches) occurred at the 0.29 threshold (i.e. when the hamming distance required for two templates to be declared a match

was 0.29 or lower)

3 errors (false matches) occurred at the 0.28 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.28 or lower)

1 error (false match) occurred at the 0.27 threshold (i.e. when the hamming distance required for two templates to be declared a match was 0.27 or lower)

Therefore, by enforcing a decision policy whereby an HD of less than 0.27 is required for a match to be declared, a 9000 person probe can be compared against a 100,000 person database with only one false match. No template comparisons resulted in an HD less than 0.26; it is uncertain how many more comparison would have been necessary for a false match to have occurred within this threshold.

Areas not Addressed in Cross-Comparison Testing

A major gap in the relevance and extensibility of Cross-Comparison Testing data to a 1:300m matching environment is the lack of false non-match rate (or false reject rate) data. In order for a test's false match rate to be meaningful, it is necessary to determine the percentage of duplicate enrollees who would have evaded detection in a 1:N search. In order to generate such data, it is necessary to acquire at least two biometric records from a given individual (one for the probe, one for the gallery). Iridian indicates that they will execute such testing shortly.

Cross-Comparison Testing also does not address enrollment rates. The data used to execute testing was gathered from operational databases consisting of enrolled individuals. The lack of reported FTE may also impact the reported false match rates; if high enrollment quality thresholds were established for these databases,

then no low-quality images would have been accepted for enrollment. This may or may not in turn impact false match or false non-match rates.

Use of Single Iris

This test used for identification only a single iris from each individual. According to Iridian's internal test data, which must be validated or reproduced independently, an individual's left and right iris differ to nearly the same degree as different individuals' irises. Therefore it is possible that utilizing both irises can provide a near-multiplicative effect on accuracy at scale (absent factors such as acquisition effort). This is in contrast to fingerprint technology, in which studies have shown that the physiology of fingerprints is such that the level of accuracy provided by N-fingerprint systems is much less than multiplicative. In other words, individuals' fingerprints are correlated, such that an individual with fingerprint characteristics *X* for a given finger has a greater than random chance than his or her other fingerprints will have characteristics similar to *X*.

Iris Recognition Decision Policies

Because a single technology suppler is responsible for nearly all iris recognition deployments, the potential for iris recognition technology to be used for 1:300m identification is closely bound to the core technology of a single firm (Iridian). As opposed to most biometric providers whose systems are deployed with variable accuracy and security setting to account for differing deployer requirements, Iridian has historically held to a policy by which its systems are deployed with a false accept rate no greater than 1 in 1.2 million. This rate applies to both 1:1 and 1:N systems, such that Iridian's technology will attempt to meet this 1 in 1.2 million false *accept* rate (measured in terms of 1:N database searches) by driving

its single-template false *match* rate (measured in terms of 1:1 matches) to extremely low levels. Regardless of database size or transaction loads, the threshold at which two templates are declared to match is automatically adjusted to enforce what Iridian has established through internal evaluation as a 1 in 1.2 million false accept rate.

The reason that this policy is notable is that on a very large-scale database, Iridian's system would be designed to automatically drive false accept rates (what we refer to as effective false match rates) to a lower level than may be required in a 300m record system. As a result, the system's false non-match rate (or false rejection rate, viewed from a 1:N perspective) may become excessively high. However, further testing is necessary to determine how the false non-match rate is impacted by such a decision policy. The following table demonstrates how Iridian enforces decision policies such that its false accept rate remains low, regardless of database size.

HD Threshold	Database Size	Observed False Matches	Estimated FMR	Estimated FAR
0.31	10^0	157	1.60×10^{-7}	1.60×10^{-7}
0.30	10^1	32	3.25×10^{-8}	3.25×10^{-7}
0.29	10^2	10	1.02×10^{-8}	1.02×10^{-6}
0.28	10^3	3	3.05×10^{-9}	3.05×10^{-6}
0.27	10^4	1	1.02×10^{-9}	1.02×10^{-5}
0.26	10^5	0	3.0×10^{-11}	3.0×10^{-6}
0.25	10^6	0	3.16×10^{-12}	3.16×10^{-6}
0.24	10^7	0	3.16×10^{-13}	3.16×10^{-6}
0.23	10^8	0	2.79×10^{-14}	2.79×10^{-6}

Figure 16: Iris Recognition Decision Policy and False Match Rates

Page 80: [5] Deleted **Michael Thieme** **7/30/2003 6:38:00 AM**

NOTE: RESPONSES FROM TECHNOLOGY PROVDERS REGARDING CURRENT DEPLOYMENTS HAVE BEEN RECEIVED IN PART. WE ANTICIPATE RESPONSES, FULL OR PARTIAL, FROM SEVERAL

ADDITIONAL PARTIES. THIS ADDITIONAL MATERIAL WILL BE PRESENTED AND ANALYZED ONCE SUFFICIENTLY REPRESENTATIVE RESPONSES ARE GATHERED FOR EACH TECHNOLOGY.

Page 101: [6] Deleted **Michael Thieme** **7/31/2003 11:09:00 PM**

“

Page 101: [6] Deleted **Michael Thieme** **7/31/2003 11:09:00 PM**

”

Page 101: [6] Deleted **Michael Thieme** **7/31/2003 11:09:00 PM**

“

Page 101: [6] Deleted **Michael Thieme** **7/31/2003 11:09:00 PM**

”

Page 101: [7] Deleted **Michael Thieme** **7/31/2003 11:19:00 PM**

db

Page 101: [7] Deleted **Michael Thieme** **7/31/2003 11:09:00 PM**

4

Page 101: [7] Deleted **Michael Thieme** **7/31/2003 11:18:00 PM**

-

Page 101: [7] Deleted **Michael Thieme** **7/31/2003 11:18:00 PM**

(span:

Page 101: [7] Deleted **Michael Thieme** **7/31/2003 11:18:00 PM**

Page 101: [7] Deleted **Michael Thieme** **7/31/2003 11:18:00 PM**

)

Page 101: [8] Deleted **Michael Thieme** **7/31/2003 11:19:00 PM**

provided herein for

Page 101: [8] Deleted **Michael Thieme** **7/31/2003 11:19:00 PM**

systems only

Page 101: [9] Deleted **Michael Thieme** **7/31/2003 11:25:00 PM**

n

Page 101: [9] Deleted **Michael Thieme** **7/31/2003 11:25:00 PM**

AAMVA

Page 101: [9] Deleted **Michael Thieme** **11/10/2003 6:49:00 AM**

n imposter

Page 101: [10] Deleted **Michael Thieme** **7/31/2003 11:22:00 PM**

Probe

Page 101: [10] Deleted **Michael Thieme** **7/31/2003 11:22:00 PM**

fails to

Page 101: [10] Deleted **Michael Thieme** **7/31/2003 11:24:00 PM**

–

Page 101: [11] Deleted **Michael Thieme** **7/31/2003 11:24:00 PM**

FMR-FNMR provided herein for best systems only

Page 101: [11] Deleted **Michael Thieme** **7/31/2003 11:26:00 PM**

o an AAMVA

Page 101: [11] Deleted **Michael Thieme** **7/31/2003 11:24:00 PM**

Page 101: [12] Deleted **Michael Thieme** **7/31/2003 11:23:00 PM**

returns rank 1

Page 101: [12] Deleted **Michael Thieme** **7/31/2003 11:22:00 PM**

-

Page 101: [13] Deleted **Michael Thieme** **11/10/2003 9:54:00 AM**

n

Page 101: [13] Deleted **Michael Thieme** **11/10/2003 6:49:00 AM**

imposter