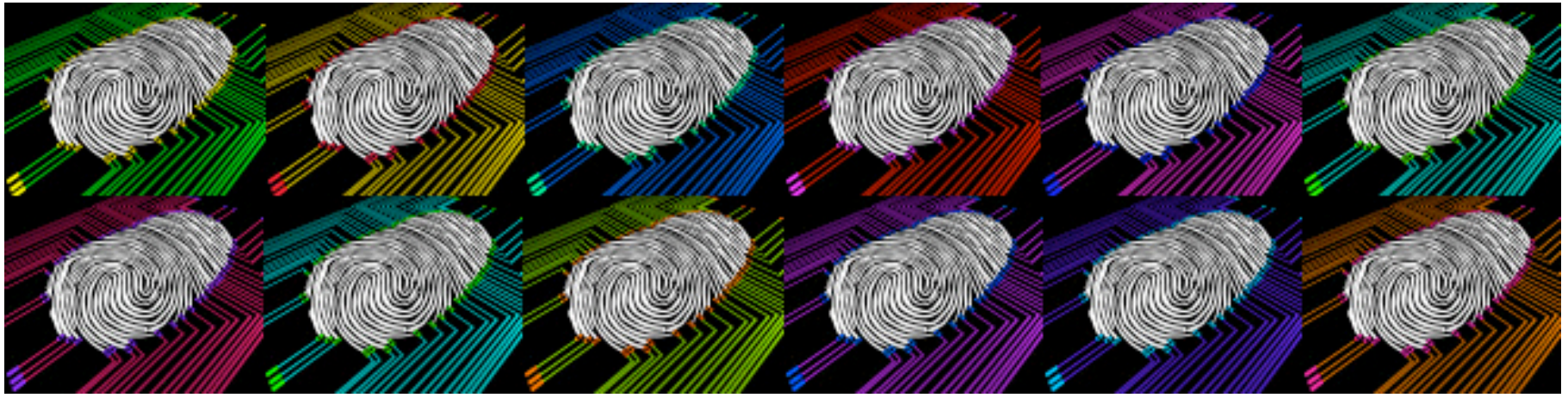
Sagem Sécurité
Groupe SAFRAN

Reducing Risk Through Large Scale Testing

Peter Waggett (IST/44 Chairman and IBM), Henry Bloomfield (Deputy Director Science and Innovation, IPS), Bill Perry (IPS), John Marc Gibbon (IBM), Jeremy Monroe (IBM), Jean-Christophe Fondeur (Sagem Sécurité)

2/3/2010



IBM

Sagem Sécurité
Groupe SAFRAN

Identity Schemes for Nation Sized Populations

- Need to provide comprehensive and complete coverage for a population - no one can be excluded
- Two major tasks:
  - Biographic analysis to establish an ‘identity’
  - Biometric analysis to associate an individual to that ‘identity’
- Exception handling processes and procedures needed

Sagem Sécurité
Groupe SAFRAN

Testing for Identity Schemes

- Identity scheme testing provides significant challenges for biometric matching due to the proposed scale and complexity of uses
- All biometric modalities to be used need to be tested at scale
- As a part of a procurement process, IBM led a team that performed performance and accuracy testing of fingerprint matching software in late 2008 and early 2009
- The lessons learnt from this testing are reported here for the fingerprint testing undertaken.
- Major facial matching testing was also performed but this will be reported at a later date

Sagem Sécurité
Groupe SAFRAN

Biometric Testing Rationale

- ID schemes using biometrics for 100M + members require significant investment in risk reduction and large scale testing activities as elements of their operation have not been fully tried and tested
- Benefits realizable from such ID schemes need to be quantified to justify expenditure on them
- Extrapolations can introduce uncertainty - large scale testing needed to minimize them

Sagem Sécurité
Groupe SAFRAN

National Identification Scheme Fundamentals

- Two main tasks for testing
  - De-Duplication - NOT identification. Does an enrollee claiming to be a new applicant of a system have a biometric reference already within a database. This requires a duplicate enrollment check and not an identification check
  - Verification. Does a supplied biometric sample match a claimed biometric reference

Sagem Sécurité
Groupe SAFRAN

Duplicate Enrollment Check

- Purpose is to determine that a biometric reference supplied as a part of an enrollment process does not match any existing biometric reference within the database. The majority of enrollments are expected to be from new applicants and the results supplied will be a negative confirmation.
- Failure to match an existing enrollment will result in a 'second' identity. This will not be flagged by the automated system and would lead to a compromised database. Prevention of this is the major task of such a system.
- Miss-identification will be flagged by such a system for further investigation. This will inconvenience the applicant and would lead to further, potentially manual investigation
- Note that the biometric matching cannot address the challenge of whether a claimed 'identity' is correct

Verification

- Confirming that a claimed identity is associated with the same biometrics from an individual who has previously enrolled.
- Impersonation needs to be made difficult at a controlled sampling station with ‘liveness’ detection
- Missed matches flagged and subject to further investigation. Will inconvenience applicant and can lead to issues in terms of wait times, queues and manning levels
- Easier task than duplicate enrollment check

Sagem Sécurité
Groupe SAFRAN

What are the Objectives for Testing?

- Customer
- IBM - Integrator
- Sagem - Sécurité - Biometric Matching Software Supplier

Customer Testing Objectives

- Overall system with lowest risk
- Quantification of business benefits
- Certainty in delivered solution
- No surprises!
- (See also talk by Marek Rejman-Greene)

Sagem Sécurité
Groupe SAFRAN

Integrator Testing Objectives

- Confirm benefits of service oriented architectures (SOA's) can be applied to biometric matching services
- Ensure competition from suppliers on a level playing field can be obtained and maintained throughout a programme
- Quantification of biometric matching performance

Biometric Matching Software Provider Objectives

- Quantification of performance obtainable from software working on representative data

Ground Rules for Integrator Testing

- 'Software' only solutions were considered. All offerings had to run on standard hardware and software platform. All solutions had to be capable of being encapsulated as a 'service' and being invoked by a service oriented architecture
- Fixed and limited time was allowed for installation, tuning and testing
- Fully automated matching results were assessed - no manual assistance or interpretation of results was included in assessment.
- Vendor selection based on task that integrator assessed as the most challenging - accuracy and performance on duplicate enrollment task
- All fingerprints in database were to be enrolled - no failures to enroll would be excluded from assessment
- No meta-data was supplied or could be used in matching task or assessment
- All work done on-site under approved secure conditions
- All data physically destroyed after testing

Sagem Sécurité
Groupe SAFRAN

Datasets Used for Fingerprint Matching Testing

- Real data
- All data supplied in a pseudo-anonymized condition with no meta-data associated with it
- 2 print (flat), 4 print (flat) and 10 print record sets (flat and rolled)
- Test data set of 100,000 individuals including 10,000 identified cases of a dual enrollment known to integrator and vendors (20,000 of supplied sets)
- Main test data sets consisting of two sets of 1.6 million records of which 300,000 in each set had a corresponding matched enrollment in the other set. Two main datasets were generated for testing purposes
  - 1 million record set with a subset of matched records known to integrator and customer only (vendor selection dataset)
  - 1 million record set with a subset of matched records known to customer only (blind dataset)

Sagem Sécurité
Groupe SAFRAN

Test Platform

- Hardware
  - 43 x IBM Intel X3550, 2 x Xeon Quad Core 3.0 Ghz with 10GB running Redhat Enterprise Linux or Suse Linux V11
  - 1 x IBM Power 5 p550 8 cores, 4Ghz with 64GB RAM running AIX 5.3.
  - IBM DS4700 Disk Array with 10 TB of storage.
- Software
  - Oracle 11g Database.
  - WebSphere Business Process Monitor
  - WebSphere MQ V6.
  - WebSphere Message Broker V6.
- Monitoring and administration
  - Xcat extreme cluster administration kit.
  - NMON performance monitor
  - Ganglia performance monitor

Sagem Sécurité
Groupe SAFRAN

Proof of Concept Platform

- Objective was to demonstrate SOA principles would work for biometric algorithms and mitigate risk
- Test Platform plus following additional components:
  - 7 x IBM Intel X3550 2 x Xeon Quad Cores 3.0Ghz with 16GB RAM running RedHat Enterprise Linux V5
  - WebSphere Business Process Monitor V6
  - WebSphere MQ V6.
  - WebSphere Message Broker V6.1.2

Sagem Sécurité
Groupe SAFRAN

Testing and Integration Enablers

- Flexible use of test platform through partitioning and procedures
- Common support and data management
- Rapid deployment of software
  - XCAT used to deploy images rapidly and repeatedly
- Oversight of testing
  - Monitoring of implementation activities
  - Performance and throughput monitoring at all levels
- Common analysis and reporting

Sagem Sécurité
Groupe SAFRAN

Testing Programme

- Set up and tuning phase => Tuned algorithms and confirmed candidates
- Vendor selection phase => Selection of vendor
- 'Blind' confirmation phase => Confirmation of results
- Verification testing => Selection of vendor
- Analysis phase => Risk mitigation
- SOA Demonstration phase => Risk mitigation

Set Up and Tuning Phase

- Biometric matching performance is affected, to some degree by a number of factors
- A major source of variation includes the characteristics of the population being matched
- Biometric Software Suppliers were provided with a limited time to ‘tune’ their algorithms. They had access to a randomly selected subset of the data with a number of known ‘matched’ enrollments within it.
- All suppliers were asked to tune their matcher thresholds to a given FMR on the test datasets
- The resulting software and configurations were frozen for all testing. No subsequent changes were allowed that would change either performance or accuracy of matching in later tests

Set Up and Tuning Phase Data

- 100,000 individuals records
  - 10,000 records which had a matching record from a second enrollment of an individual.
  - 80,000 randomly selected records
- Individuals included 2-print (flat) 4 print (flat) and 10 print sets (rolled and flat)

Sagem Sécurité
Groupe SAFRAN

Set Up and Tuning Phase

- 4 vendors established, set up, analysed performance and tuned their algorithms on the same test platform and test dataset.
- All vendors configured and provided a suitable implementation and settings for subsequent tests.
- All vendors demonstrated they provided high accuracies and performance in this phase.
- All vendors confirmed they were happy to proceed to subsequent tests with given configurations
- No detailed evaluation was made of results by the integrator or customer.

Sagem Sécurité
Groupe SAFRAN

Vendor Selection Testing

- Focussed on duplicate enrollment checking accuracy
- Performed by all 4 vendors who completed set up phase
- Essential for integrator to select optimum biometric matching software.
  - No comparable reference sites available
  - No comparable large scale testing results available
- Vendors could either enroll all data and then check entire dataset for duplicate enrollments or perform a duplicate enrollment check as each individual was enrolled
  - Different tests that required different analysis methods
  - Impacts on achievable accuracy and performance
- Quality of data and its impact on accuracy and performance was a major concern. Integrator conducted extensive analysis of data using NFIQ and MINDTCT software for subsequent analysis

Sagem Sécurité
Groupe SAFRAN

Vendor Selection Testing Dataset

- One million individuals ‘flat’ records selected for trial (91% ten print)
- These included 10,000 duplicated enrollments (known to customer and integrator but not by vendors)
- Vendors reported numbers of duplicated enrollments found
- Throughput and performance of system measured throughout testing phase

Sagem Sécurité
Groupe SAFRAN

Vendor Selection Testing Results

- Logs produced by biometric software supplier included for each record:
  - records that produced the 5 highest potential comparison match scores
  - quality scores produced by images
  - individual finger matching scores
  - whether the biometric matching software considered this a match
- Initial analysis by integrator reported number of expected matches found by comparison with ‘truth table’ and also unexpected matches reported.
- Missed matches and unexpected matches were subject to confirmation by expert manual fingerprint analysis

Vendor Selection Testing

- All vendors returned results for testing and were able to complete tests in allocated time periods
- Selection of vendor by integrator for inclusion in integrated solution was made considering:
  - Minimal residual risk of solution
  - Accuracy and performance of software
  - Total costs for solution (hardware and software)
- Sagem Sécurité selected

Sagem Sécurité
Groupe SAFRAN

Vendor Selection Testing Results

- The purpose of this phase was to discriminate between the candidate vendors to enable the integrator to select one for the inclusion in the integrated solution.
- The selection of the vendor was done on the raw results achieved against the given 'truth table' No investigation or allowance was made for data quality, failure to enroll or for manual checking of records. These results are clearly not realistic for the final analysis of system performance but provided a level against which assessment could be made.
- The following performance levels were evaluated on the test dataset during this phase
  - 'False matches'
  - 'Matches missed'
- Poor quality images were a major contribution to errors and were subject to later analysis

Sagem Sécurité
Groupe SAFRAN

‘Blind’ Confirmation Test

- 1 million records were selected randomly
- The integrator and biometric software supplier did not know how many matched enrollments were within the records (the customer did!)
- The vendor selection test process was repeated for the integrator chosen biometric software supplier
- Results were supplied to the customer for independent confirmation
- Results confirmed vendor selection test results

Sagem Sécurité
Groupe SAFRAN

Extended Analysis

- Customer, integrator and biometric software supplier conducted extensive analysis of initial results to characterise data and results
- Analysis investigated risks associated with:
  - Data quality
  - Failures to enrol
  - Manual confirmation of unexpected results
- Analysis enabled raw data results to be converted to representative data results
- Analysis was to demonstrate how risks were addressed and mitigated not to ‘prove’ final system performance

Sagem Sécurité
Groupe SAFRAN

Verification Testing

- Verification testing was performed using all 300,000 matched enrollments
- Missed match rate evaluated
- Poor quality images major contribution to errors and subject to later analysis

Sagem Sécurité
Groupe SAFRAN

Extended Analysis

- Customer, integrator and biometric software supplier conducted extensive analysis of initial results to characterise data and results
- Analysis investigated risks associated with:
  - Data quality
  - Failures to enrol
  - Manual confirmation of unexpected results
- Analysis enabled raw data results to be converted to representative data results
- Analysis was to demonstrate how risks were addressed and mitigated not to ‘prove’ final system performance

Sagem Sécurité
Groupe SAFRAN

Analysis Tasks

- Majority of analysis focussed on the risks associated with image data quality and the impact on system performance
- Traditional measures also derived (CMC, ROC curves)

Sagem Sécurité
Groupe SAFRAN

Data Quality Analysis Tasks

- Fingerprints provided from over 3 million individuals were evaluated
- NFIQ and MINDTCT scores were obtained and grouped by individual

Sagem Sécurité
Groupe SAFRAN

Determination of Suitability of Fingerprints Images from an Individual for Biometric Processing

- Individuals can have multiple instances of a biometric characteristic (e.g. up to 10 fingerprints)
- Traditional quality measures (e.g. NFIQ) give scores for one instance
- Challenge is to determine suitability of fingerprint images given by an individual for inclusion within performance measures for the system

Sagem Sécurité
Groupe SAFRAN

Derived ‘Rule’ for Determining Suitability

- Individual has at least as many fingerprint images giving a NFIQ 1-3 as there are NFIQ 5 fingerprint images.

Sagem Sécurité
Groupe SAFRAN

SOA Demonstration Phase

- Integration risks centered on need to ensure long term viability of proposed architecture
- Key issue was ability to reduce biometric matching to a simple service that could be updated and changed at will
- Demonstration showed multiple vendor algorithms running on a common platform under a controlling software ‘harness’
- Software ‘harness’ built using industry standard software products configured for application

Ongoing Research and Development Activity

- Ongoing research into service oriented architectures based on industry standards based middleware to reduce the entry barriers for new biometric matching algorithms
- Utility and capabilities of pre-processing software to improve matching of poor quality fingerprint images

Sagem Sécurité
Groupe SAFRAN

Summary and Conclusions

- Large scale testing provides a mechanism to investigate and mitigate risks associated with biometric matching
- Large scale testing as outlined in this paper can provide rapid assessment of system capabilities and design
- Large scale testing can address diverse objectives of customer, integrator and biometric software suppliers