Massive Scale Biometric Authentication System

Reimagining role of biometrics in National ID program

Raj Mashruwala & Vivek Raghavan

Think UID

Disclaimer

- Content is based on published information.
- Opinions and interpretations are of the presenter, not of Govt. of India.

Agenda

- Context
- Universal Authentication Services
- Scale & Strategy
- Challenges
- Proposed Approaches

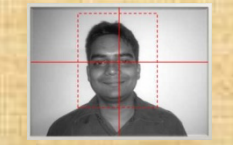
CONTEXT

Enrollment

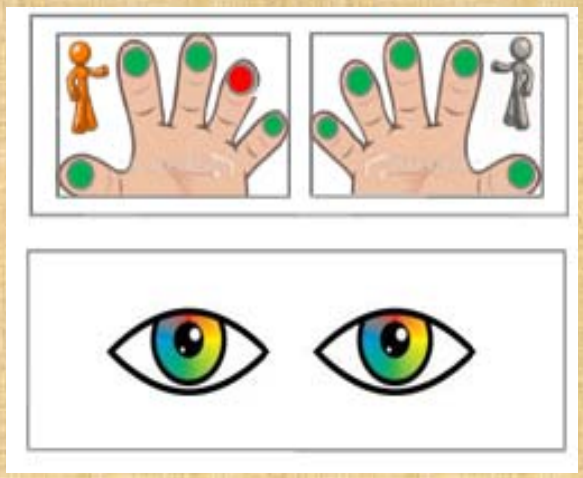
Demographic Data

- Mandatory data:
  - Name, Age/Date of Birth, Gender and
  - Address of the resident.
- Optional data:
  - Mobile number
  - Email address

Biometric Data



Photograph



All 10 Fingerprints

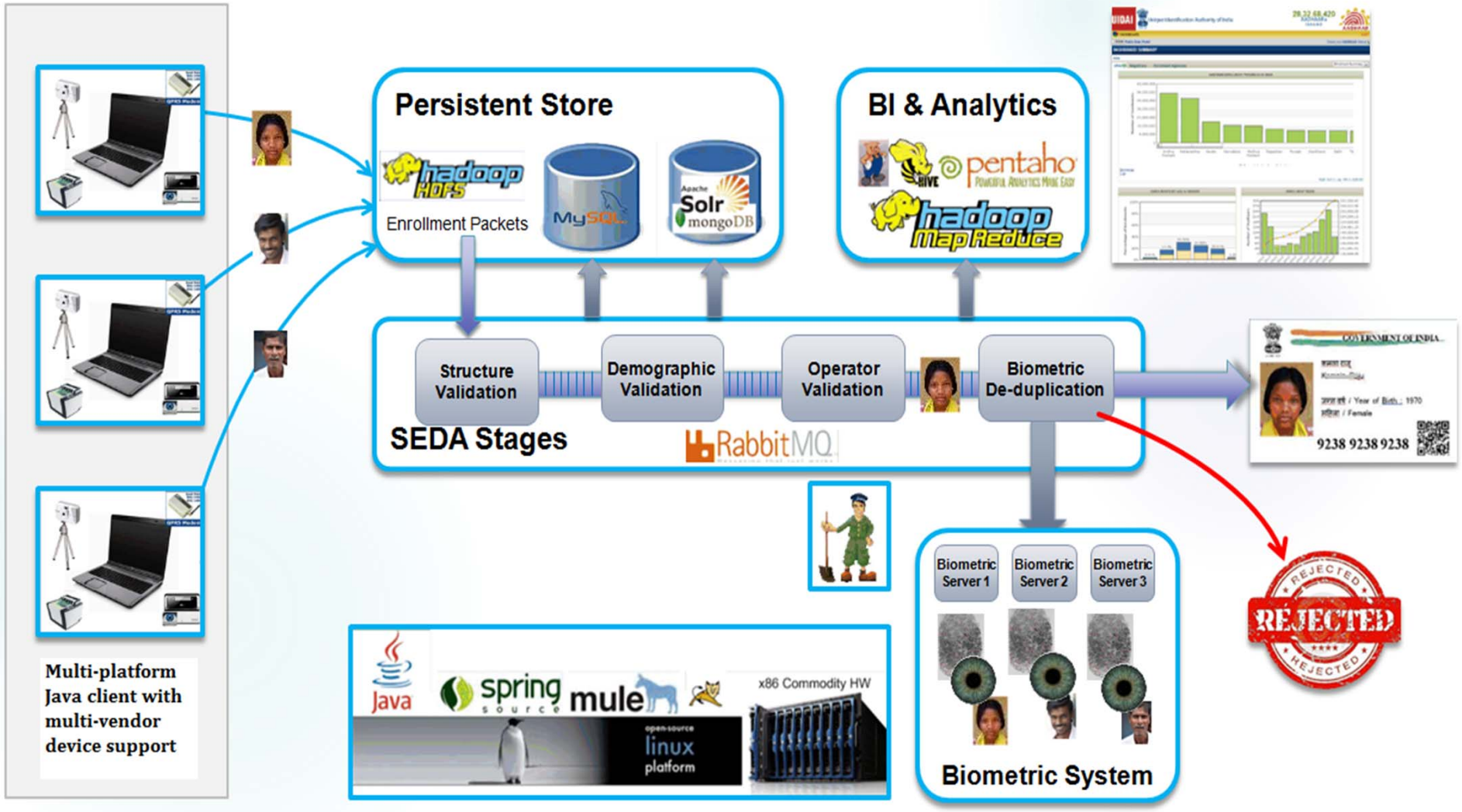
Both Iris

12-digit Aadhaar Number

Unique, lifetime, biometric based identity

Biometric SSN

Enrolment Processing



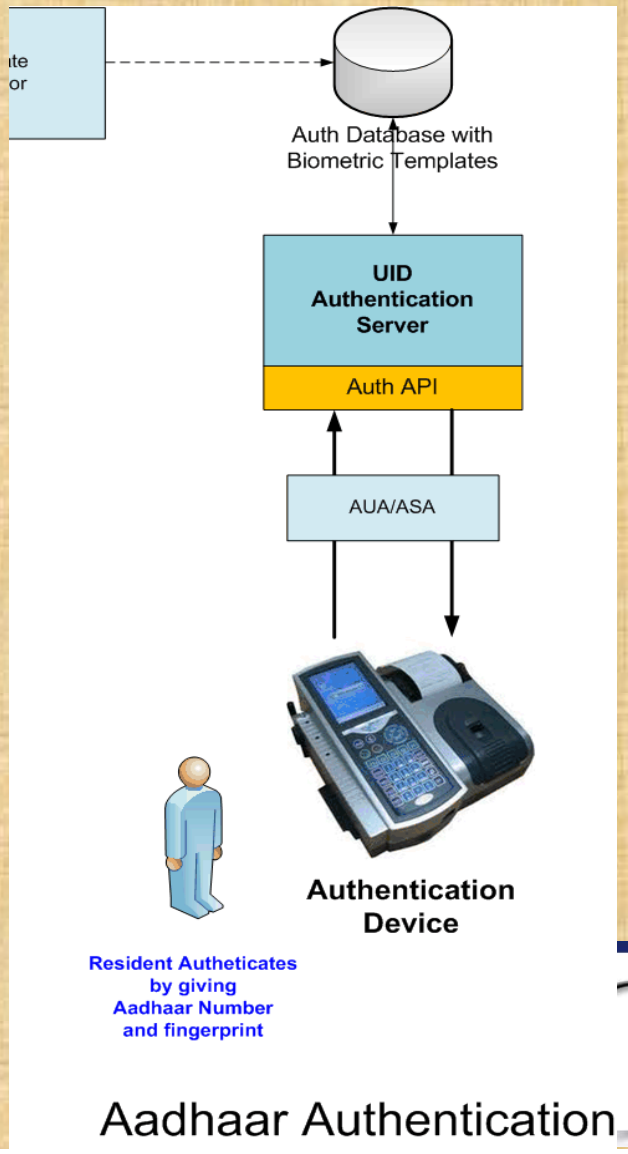
Enrolment Volume

- 600 million UIDs in 3.5 years [9/2010 until now]
  - Now processing 1 million a day
  - 400+ trillion biometric matches every day!!!
  - FNIR & FPIR ≈ 0.1%
  - 3 ABIS in parallel
- ≈ 3MB per resident packet
  - Maps to about 10 PB of raw data (2048-bit PKI encrypted!)
  - About 30 TB I/O every day
  - Replication and backup across DCs
  - Lifecycle updates and new enrolments will continue for ever
- Additional process data
  - 15+ billion records in analytics system already

AADHAAR AUTHENTICATION SERVICE

Terminology

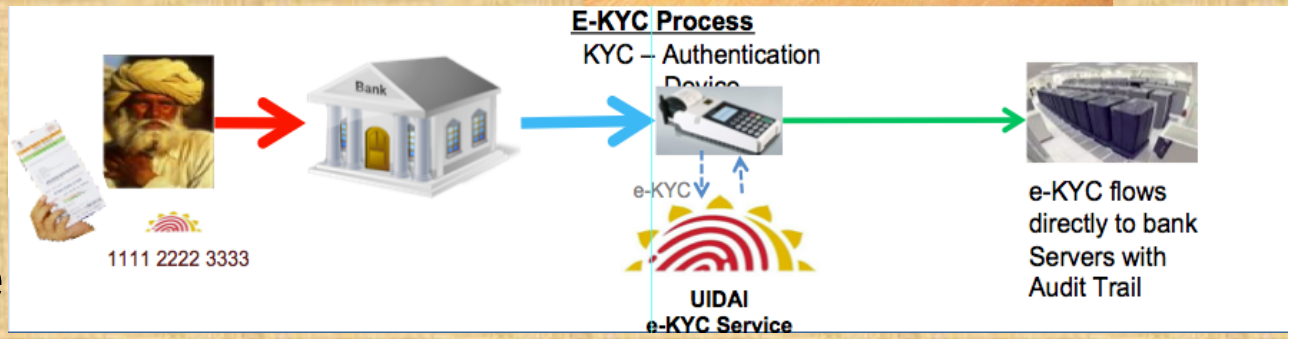
- Resident
- End point / Device
- Device model
- User agencies
- Applications



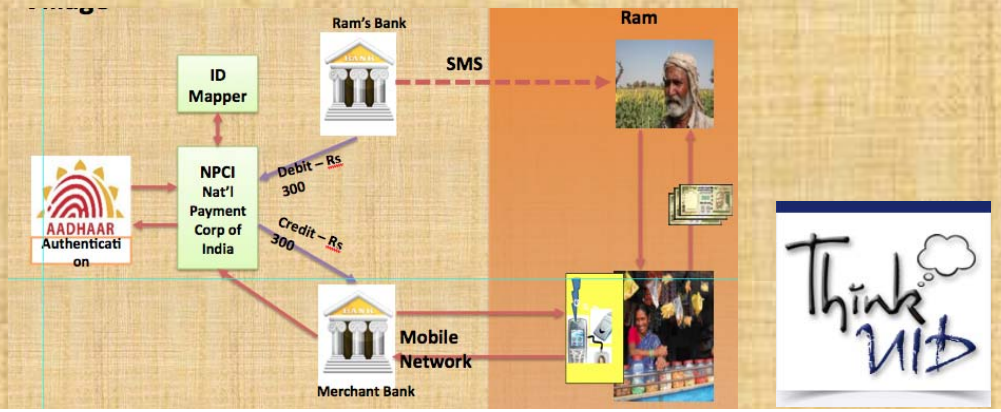
On-line Auth Uses

- Basic verification
  - Mobile network

- eKYC
  - Secure
  - Fast & free



- Financial address
  - EFT, direct deposit
  - Paypal
  - Debit card



Authentication Stats

Today

- Population covered: ≈ 50M
- Auth User Agencies: ≈ 100
- End Points: ≈ 10,000
- Daily Volume: ≈ 500,000
- Device types
  - 35 single finger sensors
  - ≈ 8 iris sensors

End Game

- 800M
- > 1000s
- > 2 million
- > 100 million
- Device types
  - ?

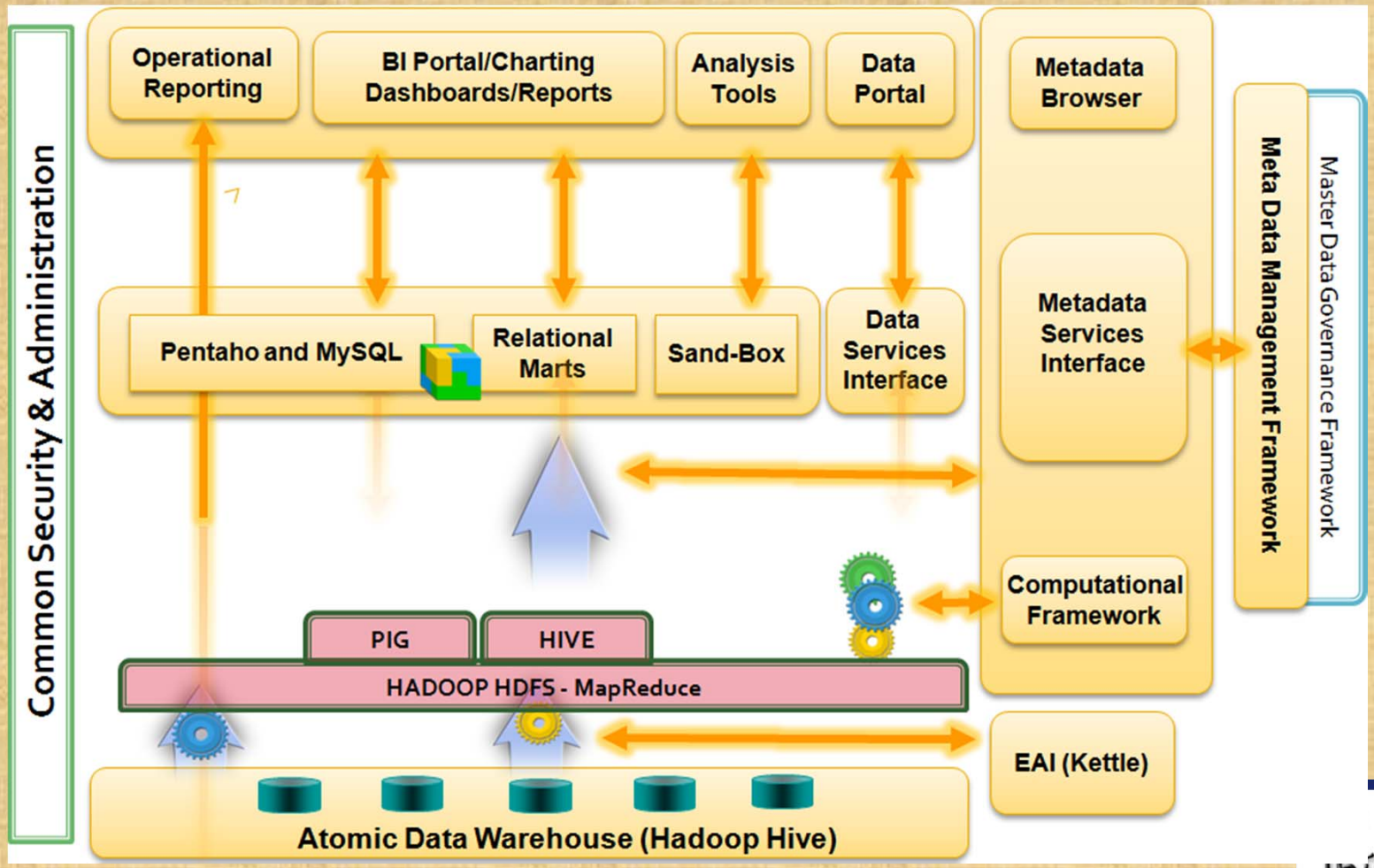
Accuracy Related Observations

- In POC,
  - FRR ≈ 2% @ FAR 1x10-4 for FP
  - FRR ≈ 0.4% @ FAR 1x10-5 for iris
- Some apps are achieving "Reject Rates" similar to the POC FRR
  - Other apps. have higher "reject rates"
- Separate "True Rejects" from "False Rejects"

Challenges of scale

- Accuracy: True Rejects vs. False Rejects
- Identification and resolution of issues
- Continuous improvement
- Security and
- Fraud Management

Approach:
Data Analytics Platform



Measurement

- Ground truth challenge
  - Measuring it in production system
  - Initialization issues/errors

Analytics around

1. True vs. false reject/accept
2. Failure to capture or coverage
3. Device performance
4. Resident experience/behavior

1. Determining Ground Truth

- Residents can be called on a sampling basis to determine:
  - False/True Accepts
  - False/True Rejects
- Can backend system identify true vs. false rejects?
  - Score just lower than threshold generally means false reject
  - Score near 0 generally means true reject

Question for Experts

- Techniques for automated estimation/ approximation
  - True Rejects vs. False Rejects
  - True Accepts vs. False Accepts

2. Coverage

- Fingerprint POC: FTC ≈ 2%
- Suspect it varies more in production but mixed with FRR/TRR.

- Factors
  - Normal factors (poor quality, device fault…)
  - Process errors: discussed above
- Image Quality Approach
  - Best Finger Authentication
  - Iris authentication
  - Non-biometric modality
  - Two factor authentication

Multi-Factor Authentication

Combine biometric factor with one time password (OTP) delivered on a registered mobile phone

- Lower threshold for biometric authentication (while providing resident “present” authentication)
- Can be used for “higher value” authentication
- Biometrics captured in two factor authentication can be used “improve” the gallery for future authentications
- Can it be used to estimate true rejects?

3. Device “Model” Performance

- Hetergenous System: Many types across many apps
- Need to provide feedback on usage and accuracy in field conditions
  - Data provided to the buyers
- Plan
  - Transparency portal to guide marketplace
  - Are there framework models/sites?

Device Level Accuracy

The “accuracy” of each end point device is measured on a daily basis

- Devices with accept rates lower than defined threshold are “highlighted”
  - Device Quality
  - Operator Training
  - Weather
  - Other Process Issue
- Trends of device “accuracy” are also measured

Question for Experts

- What techniques and methodologies could quickly highlight exceptions, outliers, out of bound cases?
  - Statistical Process Control?
  - Device Performance metrics framework?
- Device degradation modeling/detection?

4. Resident Experience/Behavior

- Measure accept rates at the resident level.
- Quality / process error during enrolment.
  - Update enrolment biometrics
  - Inherently poor biometrics
  - Determine “Best” Finger Authentication
  - Use Iris authentication or non-biometric modality
- Study of biometric “aging”

Security Features

- End-to-End Security:
  - Biometrics are encrypted using 2048 bit PKI at source and can be decrypted only in UIDAI data centers
- Locking biometrics
  - Resident can lock biometric authentication through the registered mobile
  - Unlocked just before auth. for a short time window
- Two-factor Authentication
  - Apps. may use two factor auth. for higher security
- Resident Notification
  - Each biometric authentication is notified through SMS / E-mail

Registered Devices

- Aadhaar has introduced the concept of registered devices
  - Device identification – every physical sensor device having a unique identifier allowing device authentication, traceability, analytics, and fraud management.
  - Eliminating use of stored biometrics – every biometric record is processed and encrypted within the firmware within the secure zone eliminating transmission of unencrypted biometrics from sensor to host machine.

- http://www.uidai.gov.in/images/aadhaar_registered_devices_1_0.pdf

Fraud Management

- Analytics to identify potential fraudulent transactions
  - Velocity of transaction
    - By device, by resident
  - Non-typical transaction by resident
  - Accuracy anomalies
- Call resident to determine the ground truth in suspicious circumstances

Summary

- Exciting uncharted road ahead
- Strategy: *Automated* Analytics
- Call for Help from Biometric Community
  - Measurement Techniques
  - Automation
- Interesting Results in 12 months?

Thank You