MOSIP – Deduplication Process

Overview

Deduplication is one of the core components of MOSIP which identifies and removes duplicate UIN requests based on the resident's demographic and biometric information captured during the Registration Process.

Pre-requisites

a. Registration Packet has gone through basic Packet Validation and Quality Check.
b. As configured in the registration processor workflow, the registration packet is in queue for Deduplication.

Internal Modules for Deduplication

Demographic Deduplication Module

In this module, the resident's demographic data is compared with all the processed registrations based on a pre-defined deduplication logic to get a match-ness score. The processed registrations for which match-ness scores exceeds a configured threshold are marked as duplicates.

Biometric Authentication Module

In this module, the resident's biometric data is compared with the biometric data of another resident, to ensure each of them is unique in the eco-system.

Manual Verification Module

In this module, a manual verifier verifies the data of the resident seeking UIN, with the duplicate registrations identified.
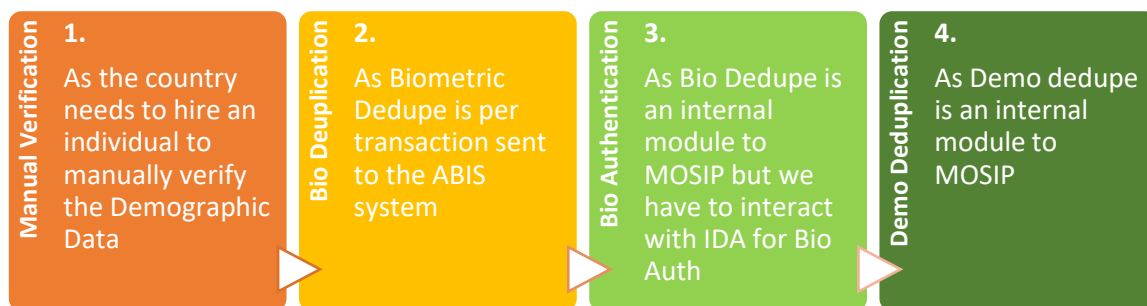
Biometric Deduplication Module

In this module, MOSIP integrates with an Automated Biometric Identification System (ABIS) to identify biometric duplicates based on a threshold configured.
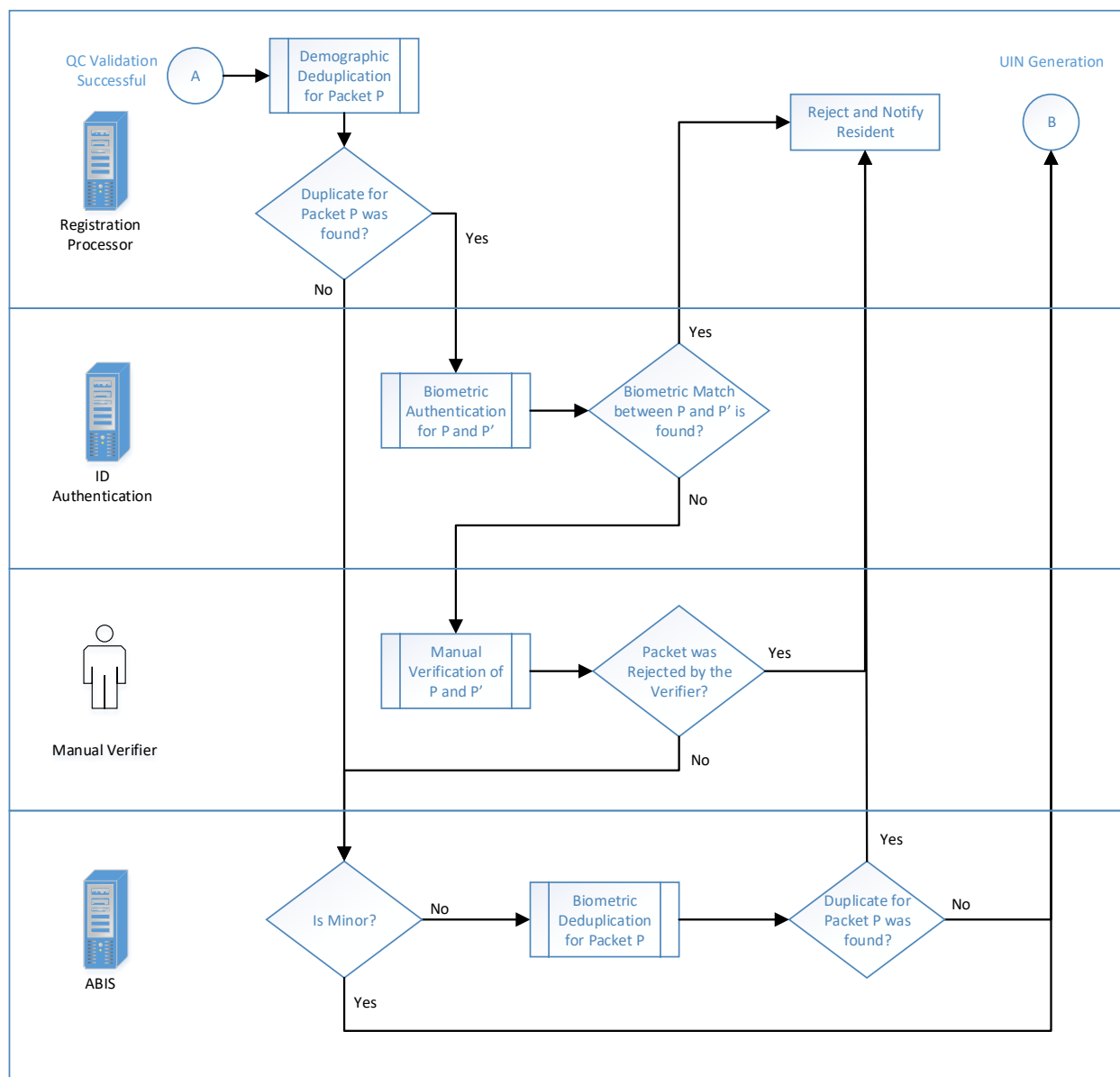
Deduplication Process Flow

Cost Analysis

We are assuming the cost associated for each module to be as follows - 1 being the most expensive and 4 being the least.

Manual Verification	Bio Deduplication	Bio Authentication	Demo Deduplication
1. As the country needs to hire an individual to manually verify the Demographic Data	**2.** As Biometric Dedupe is per transaction sent to the ABIS system	**3.** As Bio Dedupe is an internal module to MOSIP but we have to interact with IDA for Bio Auth	**4.** As Demo dedupe is an internal module to MOSIP

Hence, based on the above analysis, MOSIP recommends the below process flow:

Recommended Flow

Why Quality Check (QC) should not be Post Deduplication?

Deduplication of Packets received for UIN generation should be performed against packets for which UIN is successfully generated.

Since QC Analysis is a manual and parallel process to deduplication, some of the packets received for UIN generation will be placed On-Hold for QC Analysis (Post Deduplication).

As a result, new packets received for Deduplication will bypass the packets which are On-Hold for Quality Check.

Hence, to eradicate the possibility of duplicate registrations (by virtue of bypassing deduplication against packets on Hold for Quality Check), MOSIP recommends QC Analysis to be executed prior to deduplication; thereby ensuring unique identities are established.