• biometric identification: device specification and actual performance considered for the operations of the UIDAI

◦ *hans varghese mathews* *The Centre for Internet and Society, Bangalore*

◦ summary of findings

1 A *biometric* is a numerized representation of some generic physical feature of an organism: and the biometric devices here are those intended for the *identification of individuals.* Such a device is supplied with a *specified error rate*: the probability that the biometrics of distinct individuals will *match*.[1] When more than one device is used, and a suite of biometrics is to identify an individual, the chance of such identification errors can be derived from the specified error rates of the individual devices: and for the matching procedure the UIDAI is following we compute $0.1155001 \cdot 10^{-11}$ as *the specified identification error.*[2] The error rates specified for biometric devices would be estimated under laborotary conditions, one expects, by their manufacturers: when they are used for the rapid identification of a large population, as in our case, their performance in the field might fall short of what their specified errors promise: and we find that

• *identification error in the field considerably exceeds, by a factor of 6 almost, the specified identification error for the matching procedure followed by the UIDAI.*

We are able to draw our conclusions by examining the result of an experiment performed by the UIDAI when 84 million citizens had been registered in their biometric database. The process of obtaining and storing biometrics is termed *enrollment*; and the stored suites of biometrics are called *templates.* The experiment estimated the chance of a *false positive match*: which occurs when the suite of biometrics of a new individual, one who is not actually enrolled, happens to match *some or other* stored template. The chance of a false positive match is the conditional probability, therefore, of a match occuring *given* that the individual is not enrolled: and it is usually termed the *false reject rate.* The rate depends on the number of individuals already enrolled. Write $\rho(n)$ for the false reject rate when n individuals have been enrolled: the specified identification error is the chance, now, that the biometrics of a new individual will match *any one* given template: and if ξ is the specified identification error then

$$\rho(n) \;=\; 1 - [1 - \xi]^n$$

Now for an identification error of $(0.1155001) \cdot 10^{-11}$ and an enrolled base of 84 million the false reject rate should be $(0.97020084) \cdot 10^{-4}$ at most: but the UIDAI got an estimate of $(0.57725) \cdot 10^{-3}$ from its experiment. The bound on this rate

[1] Matching occurs when numerized representations are too similar: and similarity is usually decided by the *distance* between them falling below some *threshold,* for some appropriate measure of distance between the numerized representations the given device produces.

[2] The UIDAI is using iris scanners and fingerprint scanners: and has made their specified error rates available to researchers at the Takshashila Institute: who have made them public. The specified error for their make of iris scanner has been repored as $1/13100$; the specified error for the fingerprint scanner as $1/500$. The UIDAI has not published its matching procedure: but our investigations have led us to conclude the following: a match is taken to occur if both irises match and *any one* digit also does.

comes from the relation $1 - [1-\xi]^n \leq n \cdot \xi$, which holds for $0 < \xi < 1$ generally; from this, and from the relation between $\rho(n)$ and ξ above, one can get the bounds

$$\frac{\rho(n)}{n} \quad \leq \quad \xi \quad \leq \quad \frac{-\log[1-\rho(n)]}{n}$$

We have estimated the identification error in the field by using the UIDAI's experimental value as a reliable operational estimate of $\rho(n)$: and $(0.687400801)\cdot 10^{-11}$ is our estimate of what ξ must be in the field.

2 The false reject rate is one measure of the operational accuracy, in the field, of a suite of biometric devices. A equally important measure is its converse: the conditional probability that an individual is *not enrolled,* actually, given a match between his or her biometrics and some or other stored template. We shall term this *mistaken identification*: and our principal finding is that

- *the probability of mistaken identification rises considerably between the initial and final stages of enrollment: by a factor of 10, as it happens, between the first and last tenths of the population enrolled.*

We have proceeded here by estimating the total number of matches expected, and the number of false matches among these, for successive millions of individuals enrolled: for which we have used the lower of the bounds on $\rho(n)$ given by

$$n \cdot \xi \cdot \left[\frac{1-\xi}{1-\xi+n\cdot\xi}\right] \quad \leq \quad 1 - [1-\xi]^n \quad \leq \quad n \cdot \xi$$

The actual numbers are not negligible. The UIDAI should expect a total of $534,010$ matches to occur for the first 100 million enrolled: out of which $34,180$ will be mistaken matches. But a total of $1,280,208$ matches are expected for the last 100 million enrolled: and among these fully $780,382$ would be mistaken matches. The discrepancy is even more extreme for small initial and final subsets: we estimate $50,325$ matches for the first 10 million, of which only 341 would be mistaken ones; but $13,1050$ matches are expected for the last 10 million, out of which $8,1607$ would be mistaken matches.[3]

3 When a match occurs the UIDAI must decide whether or not the individual is already enrolled: for which the templates matching that person's suite of biometrics must be examined. The amount of work here depends on how many templates will match a given suite of biometrics, generally, when a match does occcur. We get an upper bound of $10,922,437$ on the total number of matches when the entire population of 1.2 billion has been enrolled: of which $4,924,539$ would be mistaken ones. We estimate that only $11,267,203$ matching templates will have to examined, however, to decide which matches are mistaken: and our last finding is that

- *only occasionally will more than one matching template have to be examined, when a match occurs, in order to decide whether or not that match is mistaken.*

[3] The bounds above come from Professor Nico Temme of the CWI in The Netherlands: whose freely given help we gratefully acknowledge. To counts matches and mistaken matches one needs, besides identification error, the probability that enrolled individuals will try to register again; and one needs, as well, the chance of a match for an already enrolled person. The UIDAI has conducted an experiment which allows one to estimate the latter; and it has estimated to its satisfaction the former probability as well.