Using fingerprint image quality to improve the identification performance of the U.S. Visitor and Immigrant Status Indicator Technology Program

Lawrence M. Wein*† and Manas Baveja‡

*Graduate School of Business and ‡Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305

Edited by Burton H. Singer, Princeton University, Princeton, NJ, and approved March 8, 2005 (received for review October 8, 2004)

Motivated by the difficulty of biometric systems to correctly match fingerprints with poor image quality, we formulate and solve a game-theoretic formulation of the identification problem in two settings: U.S. visa applicants are checked against a list of visa holders to detect visa fraud, and visitors entering the U.S. are checked against a watchlist of criminals and suspected terrorists. For three types of biometric strategies, we solve the game in which the U.S. Government chooses the strategy's optimal parameter values to maximize the detection probability subject to a constraint on the mean biometric processing time per legal visitor, and then the terrorist chooses the image quality to minimize the detection probability. At current inspector staffing levels at ports of entry, our model predicts that a quality-dependent two-finger strategy achieves a detection probability of 0.733, compared to 0.526 under the quality-independent two-finger strategy that is currently implemented at the U.S. border. Increasing the staffing level of inspectors offers only minor increases in the detection probability for these two strategies. Using more than two fingers to match visitors with poor image quality allows a detection probability of 0.949 under current staffing levels, but may require major changes to the current U.S. biometric program. The detection probabilities during visa application are ≈11–22% smaller than at ports of entry for all three strategies, but the same qualitative conclusions hold.

biometrics | homeland security | policy | queues | game theory

The September 11, 2001 attacks may have been prevented if some of the 18 terrorists were apprehended as they entered the U.S. (1). To rectify this situation, the multibillion dollar U.S. Visitor and Immigrant Status Indicator Technology (US-VISIT) Program (2) takes two index fingerprint images from each visa applicant and matches these prints against those of several hundred million visa holders to detect whether the new applicant already has a visa under a different identity (3). Visitors (i.e., visa-holders, as well as citizens of 27 visa-exempt countries who are visiting for >90 days or are traveling on work or student visas) to U.S. ports of entry have two new fingerprints taken, which are compared to the original prints to verify that they are who they claim to be, and are used to identify whether visitors are on a watchlist that contains known criminals and suspected terrorists. The US-VISIT Program also uses a facial image for verification (i.e., one-to-one matching); although face recognition technology has improved over the last several years, it is not a viable tool for identification (i.e., one-to-many matching) from a pool of millions (4). Here, we develop a mathematical model to compare various biometric identification strategies at visa application and at ports of entry.

Identification, which is much more difficult than verification, is performed by software systems that compute a similarity score between any two fingerprints. Upon entry to the U.S. (a similar procedure is used during visa application), a visitor is fingerprinted during primary inspection and is assigned a pair of similarity scores, one for each index finger, against each person in the watchlist. If either the left or right score exceeds a given threshold, or the sum of the left and right scores exceeds a second threshold, then the corresponding person on the watchlist is added to the visitor's candidate list. If the candidate list for a visitor is nonempty, then this visitor is further investigated during a secondary inspection. We call a visitor illegal if he is on the watchlist and legal otherwise. If a visitor is illegal, then he will have a set of fingerprints (taken at an earlier point in time), called a mate, in the watchlist. The detection probability (called the true accept rate in ref. 5) is the probability that an illegal visitor's mate is placed on the candidate list (i.e., the similarity scores between his new set of fingerprints and his mate exceeds one of the thresholds). The false positive probability (called the false accept rate in ref. 5) is the probability that a legal visitor's candidate list is nonempty. For two index fingers and for the fingerprint databases tested by the National Institute of Standards and Technology (NIST), the identification system currently used in the US-VISIT Program had a detection probability of 0.959, independent of the watchlist size, and the false positive probability increased with the size of the watchlist and was 3.1×10^{-3} when the watchlist had 6 million people (5), which is comparable in size to the actual watchlist during entry. To avoid undue congestion at ports of entry under current staffing levels, it is necessary to maintain the false positive probability at approximately this level.

The identification system used in the US-VISIT Program also computes the image quality of a fingerprint, which reflects the inherent quality of the print and operational factors such as humidity, dirt, and finger pressure. NIST has determined that the identification performance of the biometric system is highly dependent on the fingerprint image quality, with better performance resulting from good-quality images (5, 6). The present study stems from the belief that terrorist organizations can exploit the image quality-dependent performance of the biometric identification system by choosing from their large pool of potential U.S.-bound terrorists, those that have either inherently poor image quality (e.g., worn out fingers) or deliberately reduced image quality (e.g., surgery, chemicals, sandpaper). We formulate and solve a Stackelberg game (7) in which the U.S. Government chooses the parameters for a biometric identification strategy to maximize the detection probability subject to a constraint on the mean total (i.e., primary plus secondary) biometric processing time per legal visitor, and then the terrorist chooses the image quality level to minimize his detection probability. We use this optimization problem to assess three types of strategies, including the strategy currently used in the US-VISIT Program. One of these strategies requires additional primary biometric processing, which is why our constraint is in terms of

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: US-VISIT, U.S. Visitor and Immigrant Status Indicator Technology; NIST, National Institute of Standards and Technology.

†To whom correspondence should be addressed. E-mail: lwein@stanford.edu.

Table 1. A description of the three biometric strategies

Strategy	Parameters	Condition for placement on candidate list	Optimal solution	Detection probability
Current	t_1, t_2	$\cup_{j=1}^{2} (s_j > t_1) \cup (\Sigma_{j=1}^{2} s_j > t_2)$	1317, 1818	0.526
Two-finger	$t_{11}, \ldots, t_{18}, t_{21}, \ldots, t_{28}$	$\cup_{j=1}^{2}(s_j > t_{1i}) \cup (\Sigma_{j=1}^{2} s_j > t_{2i})$	4741, 3690, 3010, 2853, 3257, 1977, 1743, 1120, 5263, 4936, 4989, 4783, 3831, 2647, 2303, 1120	0.733
Multifinger	$t_1, \ldots, t_8, n_1, \ldots, n_8$	$\cup_{j=1}^{n_i} (s_j > t_i)$	1975, 1804, 1678, 1512, 1820, 1573, 1901, 1378, 2, 2, 2, 2, 3, 4, 5, 10	0.949

The first two strategies use two fingers from each visitor, whereas the multifinger strategy uses n_i fingers if the visitor has quality $i = 1, \ldots, 8$. The values of $(n_1, \ldots, n_8)$ can be based on quality information obtained at visa enrollment. The similarity scores between the visitor and each watchlist person are (s_1, s_2) for the first two strategies, and $(s_1, \ldots, s_{n_i})$ for the multifinger strategy if the visitor has quality i. The third column describes the condition on the similarity scores for placing a watchlist person onto a visitor's candidate list, where the "∪" denotes "logical or." The last two columns give the values of the parameters that solve Eqs. **1–3,** and the corresponding optimal detection probability in Eq. **1**, assuming a watchlist size of 6 million. The "Current" strategy is used in the US-VISIT Program.

the mean total biometric processing time per legal visitor rather than the false positive probability.

The Model

In our model (see *Supporting Text*, Figs. 3–7, and Tables 2–6, which are published as supporting information on the PNAS web site), each person has an associated image quality that is an integer value between 1 (highest quality) and 8 (lowest quality) (5). This random quantity is assumed to be independent and identically distributed across people, whether they are visitors or on the watchlist. This assumption ignores the fact that right and left fingers of the same person may vary in quality, operational noise may cause a specific finger to have a different image quality when retested, and people's fingers may wear out over time, causing the image quality of an illegal visitor to be worse than that of his mate on the watchlist. Although our model and analysis can be generalized so that each image, rather than each person, can possess a different quality, more detailed unpublished data from NIST would be required to parameterize the model. However, the analysis in Fig. 3, which uses data from ref. 5 and the belief that the level of operational noise can be kept low via good training and processes, suggests that this assumption holds true in the great majority of cases and can be made without affecting our qualitative conclusions.

The two fundamental building blocks of our model are the intraperson similarity scores, which quantify the match between an illegal visitor's fingerprints with the earlier pair of prints in the watchlist, and the interperson similarity scores, which compare a visitor's fingerprints with a different person's prints from the watchlist. Because each person is assumed to have a given quality level, the intraperson similarity scores are described by a family of eight probability distributions, one for each quality level. Following ref. 8, we assume that an interperson similarity score depends on the qualities of the two matched fingerprints only via the worse of the two qualities, which allows us to consider only eight interperson similarity score distributions. Guided by quality-aggregated data from NIST (9), we use gamma distributions for intraperson scores and log-normal distributions for interperson scores (see *Supporting Text* for details).

We consider the three biometric strategies described in Table 1, which are referred to as the current strategy, the two-finger strategy, and the multifinger strategy; the two-finger strategy allows quality-dependent thresholds and the multifinger strategy allows the thresholds and the number of fingers tested to vary according to quality. It takes ≈5 s per finger to take an image of a visitor's fingerprint (3, 10) and ≈2.5 s per finger to perform the software matching against the entire watchlist (10). We assume that the watchlist matching is performed in parallel with a visitor's interview at the port of entry, so that matching does not increase a visitor's processing time (10); however, the physical fingerprinting process needs to be managed by the inspector to guarantee low operational noise and to detect fraud (e.g., artificial fingerprints). In addition, we assume that 20 min are required on average to perform a secondary inspection for each false positive, regardless of the size of the candidate list. Let d_i be the detection probability if the illegal visitor has image quality i, and let f denote the overall false positive probability; these are computed in *Supporting Text* for each of the three strategies. If we let m_1 be the mean time to image a single fingerprint, m_2 be the mean secondary inspection time of positive matches, and $p(i)$ be the fraction of visitors that have quality i for $i = 1, \ldots, 8$, then the mean total biometric processing time per legal visitor is $m_1 \Sigma_{i=1}^{8} p(i)n_i + fm_2$ for the multifinger strategy, and $2m_1 + fm_2$ for the other two strategies. Substituting the values $m_1 = 5$ s, $m_2 = 20$ min, and $f = 3.1 \times 10^{-3}$, we find that the base-case value of the mean total biometric processing time per legal visitor for the two two-finger strategies is $2m_1 + 3.1 \times 10^{-3} m_2 = 10 + 3.72 = 13.72$ s at the port of entry. The optimization problem for the multifinger strategy is

$$\text{maximize minimize}_{i=1,\ldots,8}\, d_i \qquad [1]$$

$$\text{subject to } m_1 \sum_{i=1}^{8} p(i)n_i + fm_2 \leq 13.72, \qquad [2]$$

and Eq. **2** is replaced by

$$f \leq 3.1 \times 10^{-3} \qquad [3]$$

for the other two strategies. The maximization in Eq. **1** is carried out over the parameters in the second column of Table 1, assuming a watchlist of 6 million people at the port of entry.

Results

The use of quality-dependent thresholds on the similarity scores increases the detection probability from the current level of 0.526 to 0.733 (Table 1). This strategy achieves the same detection probability for all quality levels by using a low threshold ($t_{28} = 1120$) for the worst quality (the single-finger threshold t_{18} is redundant in this optimal strategy). The multifinger strategy achieves a detection probability of 0.949 by using 10 fingers for the lowest quality, and 3–5 fingers for the next three lowest quality levels.

The detection probability can be improved by increasing the mean biometric processing time per legal visitor, i.e., by increas-

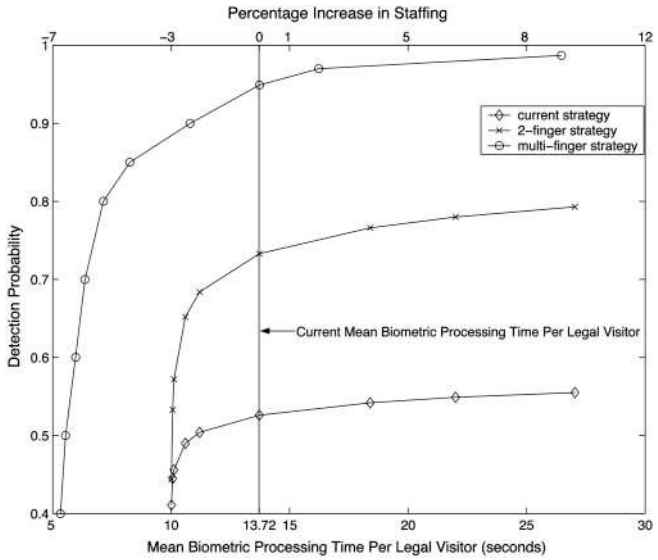

Fig. 1. Identification performance during entry to the U.S. For each of the three strategies listed in Table 1, the detection probability vs. mean biometric processing time per visitor for a fixed watchlist size of 6 million. The horizontal axis at the top of the figure measures the increase in staffing level required to process legal visitors, assuming that the mean nonbiometric processing time per visitor is 2 min.

ing the right side of Eqs. **2** and **3**. To first order (11), the mean visitor waiting time depends on the total (biometric plus nonbiometric) processing time per visitor divided by the number of U.S. Custom and Border Protection inspectors at the facility (e.g., the airport). The mean processing time per visitor in the absence of biometric inspection is ≈2 min (10) and, therefore, the increase in biometric processing time per legal visitor (we ignore the increase in staffing needed to process the additional true positives) from 13.72 to x s requires a staff increase of $\approx(120 + x - 133.72)/133.72\%$. For the two-finger strategy, a staff increase of 5% causes the detection probability to increase from 0.733 to 0.775, but further staff increases lead to decreasing marginal returns (Fig. 1). Staff increases beyond the current levels for the current strategy have a very minor effect.

During visa application, the watchlist includes all several hundred million visa holders, and widening this program's scope beyond visa-holders could increase the watchlist size to ≈1 billion within the next decade. Visitor waiting times are less of an issue during visa application because much of the processing is performed offline (10). Consequently, we fix the watchlist size to be 200 million and recompute the optimal parameter values for various values of the false positive constraint (Eq. **3**) for all three strategies (implying the optimality of $n_1 = \ldots = n_8 = 10$ for the multifinger strategy), which generates a detection probability vs. false positive probability curve (Fig. 2*a*). Fixing the false positive probability at 3.1×10^{-3}, we also compute the detection probability as a function of watchlist size (Fig. 2*b*). The increase in watchlist size from 6 million to 200 million causes a modest reduction in detection probability for all three strategies, to 0.454, 0.569, and 0.840, respectively. The difference in performance between the current and two-finger strategies increases with the false positive probability and decreases with the watchlist size. The multifinger strategy maintains a detection probability above 0.8 for watchlist sizes below ≈300M.

Discussion

Our analysis reveals that the strong dependence of biometric identification performance on image quality level (5) leaves the US-VISIT Program vulnerable to exploitation by terrorists.

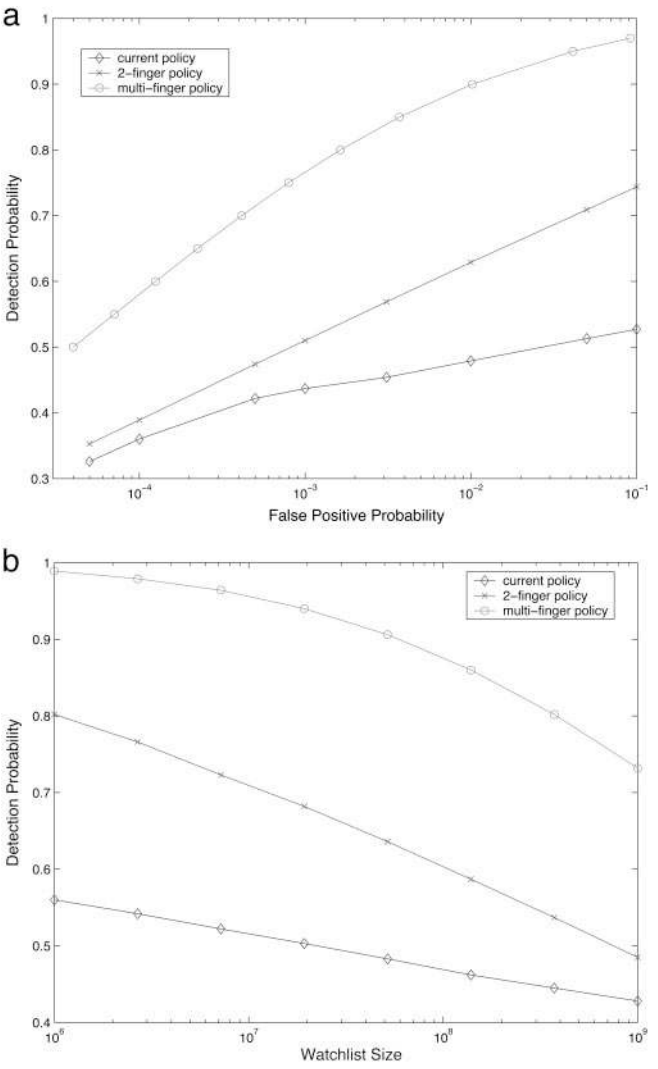

Fig. 2. Identification performance during visa application. For each of the three strategies listed in Table 1, we show the detection probability vs. the false positive probability (on a logarithmic scale) for a watchlist of size 200 million (*a*) and the detection probability vs. the watchlist size (on a logarithmic scale) for a false positive probability of 3.1×10^{-3} (*b*).

More specifically, using the existing quality-independent threshold levels, the overall quality-independent detection probability at the U.S. port of entry (0.959) is much larger than the detection probability if terrorists are allowed to game the system by choosing the worst image quality level (0.526). The detection probability can be increased from 0.526 to 0.733 by using different thresholds for each quality level. Although it is optimal for the terrorists to always choose the lowest image quality under the Government's current quality-independent strategy, in the quality-dependent scenario, the terrorists would randomize across image qualities to force the Government to equalize the detection probabilities across quality levels. Although the optimal quality-dependent two-finger strategy will likely be complex (described by a set of eight nonlinear curves in two-dimensional score space) and may offer only a modest improvement over the two-finger strategy considered here, its derivation is worth pursuing in future work.

There are several other options for improving the detection probability beyond that of the two-finger strategy. A 5% increase in the number of Custom and Border Protection inspectors can offer only a modest increase in the detection probability of the

two-finger strategy, from 0.733 to 0.775, while maintaining existing congestion levels at the ports of entry. As of May 2004, US-VISIT Program plans do not call for additional staff or facilities at land ports of entry (2). Testing 10 fingers rather than two fingers for poor-quality visitors can increase the detection probability at the U.S. border to 0.949 under current staffing levels. Moreover, during visa application, the two-finger strategy can achieve a detection probability of only 0.569 if the false positive probability is set at 3.1×10^{-3}, whereas a 10-finger strategy can achieve a detection probability of 0.840. However, the US-VISIT Program only takes images of two fingers during visa application (5), despite previous warnings that a two-finger system was inadequate for identification with large watchlists (6). Although switching from two to 10 fingers at this point in time, even for only poor-quality images, may be expensive and disruptive (6), this multifinger approach appears to be a more cost-effective exception-management alternative for poor-quality images than other biometrics (e.g., iris, retina, hand geometry; see ref. 3) or human interrogation. Slower, more accurate matching techniques for poor-quality images should also be assessed and compared to the multifinger approach. The extent to which these options are pursued should be assessed in light of the detection probability required to deter terrorists from attempting a border crossing at an official port of entry, which itself depends on the terrorists' perceived likelihood of successfully entering the U.S. between the ports of entry, e.g., along the U.S. borders with Mexico and Canada. The detection probability between the ports of entry on the U.S.–Mexico border has been estimated at 0.25 (12), although it appears that, at this point in time, Al Qaeda prefers to enter the U.S. at ports of entry (1).

The quality-dependent biometric analysis performed here has other potential applications. First, a terrorist could intentionally deface his fingerprints between the time his watchlist fingerprints were imaged and the time he enters the US-VISIT Program. If the fingerprints are only partially altered, then the low thresholds associated with poor-quality images in our two-finger strategy (Table 1) should increase the likelihood of detection slightly, and the multifinger strategy would significantly hinder the terrorists' success with this approach. This topic deserves further investigation. Another possible application is to assess the degradation in quality of faxed fingerprint images, which appears to have been at the crux of a mistaken terror arrest in Spain (13).

Although our qualitative conclusions are likely to be robust, the quality-dependent intra- and interperson similarity score distributions at the core of our model were indirectly estimated from quality-dependent detection probability vs. false positive probability curves. The use of unpublished NIST data on the intra- and interperson similarity scores broken down by image quality, and the actual watchlist, which likely generates worse performance than publicly available databases (5), perhaps coupled with a more refined model that allows quality to be associated with an image rather than with a person, would be required to sharpen our policy recommendations and to derive operationally reliable parameter values. Among other parameter values in our model, only the mean secondary processing time (m_2 in Eq. **2**) has not been reported in the literature. This parameter only affects the relative performance of the multifinger strategy, because it is the only strategy that trades off primary and secondary inspection. We may be underestimating the performance of the multifinger strategy because the four nonthumb fingers can perhaps be imaged simultaneously (14), which would reduce the mean primary processing time for a 10-finger print from 50 s to 20 s.

In conclusion, there appears to be a serious but reparable vulnerability (detection probability is highly dependent on image quality) in the biometric identification system of the US-VISIT Program, which is the last, and perhaps main, line of defense for keeping terrorists off of U.S. soil. Our analysis provides the government with a means of assessing the worst-case threat, and there is a silver lining in the equilibrium solution, namely, that the resulting detection probability will be equal for all image quality levels, rendering the system more robust than the existing system. This, in itself, is a strong reason for switching to the proposed policy. The introduction of quality-dependent thresholds requires only minor software modifications and can increase the detection probability by $\approx$0.2, and should be implemented as soon as possible. The use of more than two fingers for low-quality images can increase the detection probability to $\approx$0.95; in other words, the worst-case performance under the proposed multifinger strategy is approximately the same as the existing strategy's performance under the naive assumption that terrorists do not behave strategically at all. There is no excuse for a multibillion dollar program to settle for performance below the level of the proposed multifinger strategy, particularly given the potentially grave consequences of a false negative. Our policy recommendations hinge on the assumption that terrorist organizations will attempt to defeat the biometric system by employing terrorists with poor-quality fingerprints. In light of the meticulous planning that has gone into terrorist attacks over the last decade (1), we believe this assumption is not only prudent, but realistic.

We thank Michael Garris for sharing data from refs. 5 and 9. This research was supported by the Center for Social Innovation, Graduate School of Business, Stanford University, and by a fellowship from the Center for International Security and Cooperation, Stanford University.

1. National Commission on Terrorist Attacks (2004) *The 9/11 Commission Report* (Norton, New York).
2. U.S. General Accounting Office (2004) *First Phase of Visitor and Immigration Status Program operating, but improvements needed* (Government Accountability Office, Washington, DC), report GAO-04-586.
3. U.S. General Accounting Office (2002) *Using Biometrics for Border Security* (Government Accountability Office, Washington, DC), report GAO-03-174.
4. Phillips, P. J., Grother, P., Michaels, R. J., Blackburn, D. M., Tabassi, E. & Bone, J. M. (2003) *Face Recognition Vendor Test 2002* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology International Report 6965.
5. Wilson, C. L., Garris, M. D. & Watson, C. I. (2004) *Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology International Report 7110.
6. The Attorney General, Secretary of State & the National Institute of Standards and Technology (2003) *Report to the Congress: Use of Technology Standards and Interoperable Databases with Machine-Readable, Tamper-Resistant Travel Documents* (Office of the Attorney General, U.S. Department of State, and National Institute of Standard and Technology, Gaithersburg, MD).
7. Gibbons, R. (1992) *Game Theory for Applied Economists* (Princeton Univ. Press, Princeton).
8. Tabassi, E., Wilson, C. L. & Watson, C. I. (2004) *Fingerprint Image Quality* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology International Report 7151.
9. Wilson, C. L., Watson, C. I., Garris, M. D. & Hicklin, A. (2003) *Studies of Fingerprint Matching Using the NIST Verification Test Bed (VTB)* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology International Report 7020.
10. Edmunds, T., Sholl, P., Yao, Y., Gansemer, J., Cantwell, E., Prosnitz, D., Rosenberg, P. & Norton, G. (2004) *Simulation Analysis of Inspections of International Travelers at Los Angeles International Airport for US-VISIT* (Lawrence Livermore National Laboratory, Livermore, CA).
11. Halfin, S. & Whitt, W. (1981) *Oper. Res.* **29,** 567–588.
12. Bartlett, D. L. & Steele, J. B. (September 20, 2004) *Time*, Vol. 164, No. 12, p. 51.
13. Kershaw, S. (June 5, 2004) *N.Y. Times*, Section A, p. 1.
14. Hicklin, R. A. & Reedy, C. L. (2002) *Implications of the IDENT/IAFIS Image Quality Study for Visa Fingerprint Processing* (Mitretek Systems, Falls Church, VA), www.mitretek.org/publications/biometrics/NIST-IQS.pdf.

Supporting Information

This supporting information provides a detailed derivation of the results reported in the main text. Probability distributions for image quality, intraperson similarity scores and interperson similarity scores are described in *Image Quality*, *Intraperson Similarity Scores* and *Interperson Similarity Scores* respectively. The detection probabilities $d_1, \ldots, d_8$ and the false positive probability f appearing in Eqs. 1-3 in the main text are derived in *Detection Probability* and *False Positive Probability*. The solution to Eqs. 1-3 in the main text is described in *Solving the Optimization Problem*.

Image Quality

In our model, each U.S. visitor and each person in the watchlist is assumed to have an image quality associated with him that is an independent and identically distributed integer-valued random variable denoted by Q, where $Q = 1$ represents the best quality and $Q = 8$ denotes the worst quality. We denote $\mathbf{P}(Q = i)$ by $p(i)$ for $i = 1, \ldots, 8$.

To motivate this assumption, we refer to Table 2 and Fig. 3, which contain the probability mass function pmf for the image quality of 274k pairs of right index fingers (i.e., two different right index images are taken from each individual), and the pmf for 274k pairs of left index fingers, denoted by $p^r(i)$ and $p^l(i)$ respectively. Table 2 and Fig. 3 also report the pmf of the "search quality" $p^s(i)$, which first computes the worse quality within the left pair of each individual and the worse quality within the right pair of each individual, and then is defined as the better of these two worse qualities. These three curves are taken from figure 11 of ref. 1. If each person has an associated quality, as assumed in our model, then these three probability mass functions would be identical, and indeed it can be seen that they are very similar. However, given the limited amount of data, we cannot eliminate the possibility that there is significant variability within each pair, although this seems unlikely from a physical standpoint.

To further investigate this assumption, suppose there was no environmental noise or machine error, so that there is no variability within each left pair and within each right pair of images. If we assume that the quality of a person's left finger and right finger are independent instead of perfectly dependent, then the pmf of the search quality would be

$$p_s(i) = \sum_{j=i}^{8} p^l(i)p^r(j) + \sum_{j=i+1}^{8} p^l(j)p^r(i), \qquad 1$$

which is given in Fig. 3 and in the last row of Table 2. The actual search quality and the search quality computed under the independence assumption are quite different from each other, which provides further credence in our assumption. More specifically, the mean left index finger quality is 3.2, right index finger quality is 2.8, actual search quality is 3.0 and computed search quality is 1.9. For concreteness, we estimate $p(i)$ by the right finger pmf $p^r(i)$ in Table 2.

Intraperson Similarity Scores

By our assumption in *Image Quality*, if a visitor is on the watchlist, then his new fingerprint images (i.e., taken upon entry) and his corresponding mates in the watchlist all have the same quality. Let X^l and X^r denote the random variables quantifying the intraperson similarity scores for left and right index fingers, respectively. We define a family of eight intraperson distributions with cumulative distribution function cdf $G_i(x)$ and probability density function pdf $g_i(x)$, one for each quality level, and assume that left and right similarity scores are conditionally independent given the person's quality:

$$\mathbf{P}(X^l \leq x_1, X^r \leq x_2 | Q = i) = G_i(x_1)G_i(x_2). \qquad 2$$

That is, the dependence between the left and right similarity scores is due to these images having the same quality. In a similar manner, in Eq. 7 in *Interperson Similarity Scores* we assume conditional independence of interperson similarity scores given the quality levels of the people being matched.

Because raw similarity scores are not published in ref. 1, we use the similarity scores from section 5.13 in ref. 2, which do not have image quality information associated with them, to choose the form of the probability distribution. Fig. 4 plots all the intraperson scores along with the best-fit log-normal, Weibull, and gamma distributions. We use the gamma distribution, which appears to be as good a fit as the other two distributions, and define

$$g_i(x) = \frac{1}{\theta_i^{\lambda_i}\Gamma(\lambda_i)} x^{\lambda_i - 1} e^{-x/\theta_i}, \tag{3}$$

where λ_i is the shape parameter and θ_i is the scale parameter.

We use figure 12 of ref. 1 to derive the two parameters for each of the eight gamma distributions. This figure gives detection probability vs. false positive probability curves for each of eight quality levels. These curves contain 12 points each, which are generated by varying the pair of threshold levels corresponding to the quality-independent threshold strategy that places a watchlist person on the candidate list if the left or right similarity score exceeds t_1, or if their sum exceeds t_2. However, only two of the 12 threshold pairs are given in ref. 1. In particular, the seventh point from the right in each of these eight curves corresponds to the thresholds $(t_1, t_2) = (1300, 1880)$ and the ninth point from the right corresponds to the thresholds $(t_1, t_2) = (1400, 2025)$. These detection probabilities are provided in Table 3.

The 16 gamma parameters can be estimated from the 16 values in Table 3 by computing d_i, which is the probability of detecting a visiting terrorist with quality i:

$$d_i = 1 - \mathbf{P}(X^l \le t_1, X^r \le t_1, X^l + X^r \le t_2 | Q = i), \tag{4}$$

$$= 1 - \int_0^{t_1} \mathbf{P}(X^l \le t_1, X^l \le t_2 - x | Q = i) g_i(x) dx \quad \text{by conditional independence}, \tag{5}$$

$$= 1 - G_i(t_1) G_i(t_2 - t_1) - \int_{t_2 - t_1}^{t_1} G_i(t_2 - x) g_i(x) dx. \tag{6}$$

Using Eqs. 3 and 6 in conjunction with Table 3, we solve the 16 equations for the 16 unknowns and get the parameters stated in Table 4. These eight distributions are plotted in Fig. 5. As expected, the worse the image quality, the higher the likelihood of generating low similarity scores.

Interperson Similarity Scores

Let Y_k^l and Y_k^r be the similarity scores of the left and right index fingers, respectively, when a visitor is matched against the k^{th} non-mate person on the watchlist. In contrast to the intraperson case in *Intraperson Similarity Scores*, the visitor and the watchlist person may have different qualities. On page 10 of ref. 3, NIST reports that in matching two fingerprints, it is the worse of the two image qualities that drives the similarity score. Let the visitor have quality Q_v and let the k^{th} non-mate person on the watchlist have quality Q_k. Using NIST's observation, we assume that the similarity score between the visitor and the k^{th} watchlist person depends on Q_v and Q_k only via the worse of the two qualities, which is $\max(Q_v, Q_k)$. We model the interperson scores using a family of eight distributions with cdf $H_i(y)$ and pdf $h_i(y)$ for $i = 1, \ldots, 8$. Assuming conditional independence of similarity scores given quality, we have

$$\mathbf{P}(Y_k^l \leq y_1, Y_k^r \leq y_2 | \max(Q_v, Q_k) = i) = H_i(y_1)H_i(y_2). \qquad 7$$

As in *Intraperson Similarity Scores*, we use the quality-aggregated similarity scores in section 5.13 of ref. 2 to specify a probability distribution. For lack of quality-segregated data, we model the interperson score distributions as log-normal (see Fig. 6):

$$h_i(y) = \frac{1}{\sigma_i \sqrt{2\pi} y} e^{-(\ln y - \mu_i)^2 / 2\sigma_i^2}, \quad i = 1, \ldots, 8, \qquad 8$$

where μ_i is the mean and σ_i is the standard deviation of the underlying normal distribution.

To test the robustness of the specification of the form of the distribution, we also fit the right portion of the interperson score distributions by exponential tails (see Fig. 6), and recomputed the results (i.e., the last two columns of Table 1) for several simpler strategies that do not require the sum of the left and right scores (so that only the right tail of the interperson score distribution needs to be specified). The results for the exponential tail were very similar (detection probability within 0.02-0.03) to the corresponding log-normal results (data not shown).

To estimate the 16 log-normal parameters in Eq. 8, we again use figure 12 of ref. 1. This figure gives the false positive probability at different pairs of threshold levels for each of the eight quality levels. As stated in the previous section, the seventh point from the right in each of these eight curves corresponds to the threshold pair $(t_1, t_2) = (1300, 1880)$ and the ninth point from the right corresponds to the threshold pair $(t_1, t_2) = (1400, 2025)$. These false positive probabilities at these two threshold pairs are given in Table 5.

We now derive an expression for the false positive probability given that $Q_v = i$, which we denote by f_i. We define the event E_k to be $(Y_k^l \leq t_1) \cap (Y_k^r \leq t_1) \cap (Y_k^l + Y_k^r \leq t_2)$, where t_1 and t_2 are the one-finger and two-finger thresholds in ref. 1. If we let event E be $\cap_{k=1}^n E_k$, where n is the watchlist size, then

$$f_i = 1 - \mathbf{P}(E|Q_v = i), \tag{9}$$

$$= 1 - \prod_{k=1}^{n} \mathbf{P}(E_k|Q_v = i) \quad \text{by conditional independence.} \tag{10}$$

We have that

$$\mathbf{P}(E_k|Q_v = i) = \sum_{j=1}^{8} \mathbf{P}\Big(E_k|Q_v = i, Q_k = j)p(j), \tag{11}$$

$$= \sum_{j=1}^{8} \mathbf{P}\Big(Y_k^l \leq t_1, Y_k^r \leq t_1, Y_k^l + Y_k^r \leq t_2 | \max(Q_v, Q_k) = \max(i, j)\Big)p(j), \tag{12}$$

$$= \sum_{j=1}^{8} \Big(H_{\max(i,j)}(t_1) H_{\max(i,j)}(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_{\max(i,j)}(t_2 - y) h_{\max(i,j)}(y) dy\Big)p(j) \quad \text{by Eqs. 5 and 6,} \tag{13}$$

$$= \Big(H_i(t_1)H_i(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_i(t_2 - y)h_i(y)dy\Big)\Big(\sum_{j=1}^{i} p(j)\Big) + \sum_{j=i+1}^{8} \Big(H_j(t_1)H_j(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_j(t_2 - y)h_j(y)dy\Big)p(j). \tag{14}$$

Substituting Eq. 14 into Eq. 10 yields

$$f_i = 1 - \Bigg(\Big(H_i(t_1) H_i(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_i(t_2 - y) h_i(y) dy \Big) \Big(\sum_{j=1}^{i} p(j) \Big) + \sum_{j=i+1}^{8} \Big(H_j(t_1) H_j(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_j(t_2 - y) h_j(y) dy \Big) p(j) \Bigg)^n . \quad (15)$$

Eq. 15 and the data in Table 5 enable us to determine the 16 log-normal parameters. These parameter values are given in Table 6 and the eight distributions are plotted in Fig. 7. In contrast to the intraperson score distributions in Fig. 5, the interperson score distributions in Fig. 7 are not strictly monotonic as a function of quality. Although higher-quality distributions typically have a higher mean but a thinner right tail, quality 5 actually has the lowest mean and the fattest right tail. This lack of monotonicity is not necessarily unexpected, and may reflect the complex relationship between image quality and the underlying algorithm that computes the similarity score.

Detection Probability

For the three strategies described in Table 1, we derive here the detection probability d_i for an illegal visitor of quality i in terms of the intraperson similarity score distributions. We change notation slightly and let $X^{(j)}$ be the similarity scores when this visitor's new fingerprints are matched against his mate in the watchlist, where $j = 1, 2$ for the first two strategies in Table 1, and $j = 1, \ldots, n_i$ for the multifinger strategy (by default $j = 1, 2$ correspond to the left and right index fingers respectively). The similarity score s_j in Table 1 can be viewed as a realization of the random variable $X^{(j)}$. Similarly, for the two-finger and multifinger strategies, the strategy parameters (i.e., t_{1i}, t_{2i}, t_i and n_i in Table 1) for a visitor are random variables that are realized when the quality of the visitor (and the quality of the watchlist person when computing the false positive probabilities) becomes known. In Eqs. 18, 20, 25, and 27 below, we refer to these random variables as T_1, T_2, T, and N, respectively.

The current strategy. By Eq. 6, the detection probability is

$$d_i = 1 - \mathbf{P}\Big((X^1 \ \leq t_1) \cap (X^2 \ \leq t_1) \cap (X^1 \ + X^2 \ \leq t_2) | Q = i\Big), \tag{16}$$

$$= 1 - G_i(t_1)G_i(t_2 - t_1) - \int_{t_2 - t_1}^{t_1} G_i(t_2 - x)g_i(x)dx. \tag{17}$$

The two-finger strategy. The detection probability is

$$d_i = 1 - \mathbf{P}\Big((X^1 \ \leq T_1) \cap (X^2 \ \leq T_1) \cap (X^1 \ + X^2 \ \leq T_2) | Q = i\Big), \tag{18}$$

$$= 1 - G_i(t_{1i})G_i(t_{2i} - t_{1i}) - \int_{t_{2i} - t_{1i}}^{t_{1i}} G_i(t_{2i} - x)g_i(x)dx. \tag{19}$$

The multifinger strategy. The detection probability is

$$d_i = 1 - \mathbf{P}\Big(\cap_{j=1}^{N}(X^{(j)} \leq T) | Q = i\Big), \tag{20}$$

$$= 1 - G_i^{n_i}(t_i). \tag{21}$$

False Positive Probability

In this section we derive the false positive probability f_i for a legal visitor with quality i in terms of the interperson similarity score distributions, for the three strategies in Table 1. The overall false positive probability f in Eqs. 2 and 3 in the main text can be derived from f_i via

$$f = \sum_{i=1}^{8} f_i p(i), \tag{22}$$

where $p(i)$ is the fraction of visitors with quality i. Let $Y_k^{(j)}$ be the random similarity score for finger j when this visitor is matched against the k^{th} non-mate person on the watchlist, where $j = 1, 2$ for the first two strategies in Table 1 and $j = 1, \ldots, n_i$ for the multifinger strategy. The similarity score s_j in Table 1 represents a realization of the random variable $Y_k^{(j)}$.

The current strategy. By Eq. 15, the false positive probability is

$$f_i = 1 - \prod_{k=1}^{n} \sum_{j=1}^{8} \mathbf{P}\Big((Y_k^1 \ \ \leq t_1) \cap (Y_k^2 \ \ \leq t_1) \cap (Y_k^1 \ + Y_k^2 \ \leq t_2) | Q_v = i, Q_k = j\Big) p(j), \quad 23$$

$$= 1 - \Bigg(\Big(H_i(t_1) H_i(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_i(t_2 - y) h_i(y) dy \Big) \Big(\sum_{j=1}^{i} p(j) \Big) + \sum_{j=i+1}^{8} \Big(H_j(t_1) H_j(t_2 - t_1) + \int_{t_2 - t_1}^{t_1} H_j(t_2 - y) h_j(y) dy \Big) p(j) \Bigg)^n . \quad 24$$

The two-finger strategy. The false positive probability is

$$f_i = 1 - \prod_{k=1}^{n} \sum_{j=1}^{8} \mathbf{P}\Big((Y_k^1 \ \ \leq T_1) \cap (Y_k^2 \ \ \leq T_1) \cap (Y_k^1 \ + Y_k^2 \ \leq T_2) | Q_v = i, Q_k = j\Big) p(j), \quad 25$$

$$= 1 - \Bigg(\Big(H_i(t_{1i}) H_i(t_{2i} - t_{1i}) + \int_{t_{2i} - t_{1i}}^{t_{1i}} H_i(t_{2i} - y) h_i(y) dy \Big) \Big(\sum_{j=1}^{i} p(j) \Big) + \sum_{j=i+1}^{8} \Big(H_j(t_{1j}) H_j(t_{2j} - t_{1j}) + \int_{t_{2j} - t_{1j}}^{t_{1j}} H_j(t_{2j} - y) h_j(y) dy \Big) p(j) \Bigg)^n . \quad 26$$

The multifinger strategy. The false positive probability is

$$f_i = 1 - \prod_{k=1}^{n} \sum_{j=1}^{8} \mathbf{P}\Big(\cap_{j=1}^{N} (Y_k^{(j)} \leq T) | Q_v = i, Q_k = j \Big) p(j), \quad 27$$

$$= 1 - \Bigg(\Big(H_i^{n_i}(t_i) \Big) \Big(\sum_{j=1}^{i} p(j) \Big) + \Big(\sum_{j=i+1}^{8} H_j^{n_i}(t_j) p(j) \Big) \Bigg)^n . \quad 28$$

Solving the Optimization Problem

In this section we briefly discuss how the optimization problem presented in Eqs. 1-3 of the main text are solved. Because both the quality-dependent detection probabilities d_i computed in *Detection Probability* and the quality-dependent false positive probabilities f_i computed in *False Positive Probability* are monotonically decreasing in the one-finger and two-finger thresholds, it follows that the inequality constraints in Eqs. 2 and 3 in the main text can be replaced by equality constraints.

The current strategy. For the eight intraperson gamma distributions derived in *Intraperson Similarity Scores*, we have $d_8(t_1, t_2) < d_i(t_1, t_2)$ for all t_1, t_2 and for $i = 1, \ldots, 7$. Therefore, the optimization

problem simplifies to

$$\underset{t_1,t_2}{\text{maximize}} \quad d_8 \tag{29}$$

$$\text{subject to} \quad f = \alpha, \tag{30}$$

where α is the maximum allowable false positive probability. This optimization problem is solved using the fmincon function in matlab.

The two-finger strategy. The monotonic behavior described at the beginning of this section implies that all the d_i values will be equal to one another in the optimal solution. Although we do not give a formal proof here, suppose there was one d_i value higher than the others and one lower than the others. In this case, we could increase the thresholds corresponding to the higher d_i value, which would reduce this d_i and make the false positive constraint slack. This slackness could then be filled by lowering the thresholds corresponding to the lower d_i value, which in turn raises this lower value, thereby increasing the objective function value.

Consequently, the optimization problem can be reformulated as a maximization problem by adding a scalar variable, which we denote by β:

$$\underset{t_{11},\ldots,t_{18},t_{21},\ldots,t_{28},\beta}{\text{maximize}} \quad \beta \tag{31}$$

$$\text{subject to} \quad d_i = \beta, \;\; i = 1, \ldots, 8, \tag{32}$$

$$f = \alpha, \tag{33}$$

which is also solved using the fmincon function in matlab.

The multifinger strategy. As in the two-finger strategy, it can again be argued that all the d_i's will be equal in the optimal solution. Hence, we can reformulate Eqs. 1 and 2 in the main text as

$$\underset{t_1,\ldots,t_8,n_1,\ldots,n_8,\beta}{\text{maximize}} \quad \beta \tag{34}$$

$$\text{subject to} \quad d_i = \beta, \;\; i = 1, \ldots, 8, \tag{35}$$

$$m_1 \sum_{i=1}^{8} p(i) n_i + f m_2 = \gamma, \tag{36}$$

where γ is the maximum allowable mean total biometric processing time per legal visitor.

We propose the following greedy algorithm for this problem, where $\mathbf{n} = (n_1, \ldots, n_8)$ is the eight-dimensional finger vector that specifies the number of fingers tested for visitors of each quality, and $\mathbf{e_i}$ is the i^{th} unit vector, which has a 1 in the i^{th} component and zeroes elsewhere. We call the vector $\mathbf{n}$ feasible if $n_i \in \{1, 2, \ldots, 10\}$ for $i = 1, \ldots, 8$.

Step 1: Fix β.

Step 2: Initialize the finger vector to be $\mathbf{n} = (1, \ldots, 1)$.

Step 3: For each of the nine vectors $\mathbf{n}$ and $\mathbf{n} + \mathbf{e_i}, i = 1, \ldots, 8$ that are feasible, solve Eq. 35 uniquely for $(t_1, \ldots, t_8)$. Use these values to compute the left side of Eq. 36, which represents the mean total biometric processing time per legal visitor for each of the (up to nine) feasible finger vectors. Let $\hat{\mathbf{n}}$ be the finger vector that achieves the smallest value of the left side of Eq. 36 among the feasible finger vectors.

Step 4: If $\hat{\mathbf{n}} = \mathbf{n}$, i.e., the vector $\mathbf{n}$ achieves a smaller mean total biometric processing time per legal visitor than any of the other feasible finger vectors, then stop, and let the optimal finger vector be $\hat{\mathbf{n}}$ and the optimal thresholds be the corresponding $(t_1, \ldots, t_8)$ computed in Step 3. Otherwise, go to Step 5.

Step 5: Set the new value of $\mathbf{n}$ to be $\hat{\mathbf{n}}$, and go back to Step 3.

Note that this 5-step procedure provides a solution to the dual of Eqs. 34-36, i.e., it minimizes the mean total biometric processing time legal visitor for a specified detection probability. Repeating the 5-step procedure for various values of the detection probability (using a standard one-dimensional search) until the resulting mean total biometric processing time per legal visitor equals 13.72 seconds eventually generates a solution to Eqs. 1 and 2 in the main text. Although this greedy procedure only guarantees a local optimum to the mixed integer and continuous program described by Eqs. 1 and 2 in the main text, the monotonic behavior of the d_i and f_i functions and the relation between the various intraperson and interperson similarity score distributions suggest that this algorithm indeed identifies

the globally-optimal solution.

References

[1] Wilson C. L., Garris M. D. & Watson C. I. 2004 *Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology Internal Report 7110.

[2] Wilson C. L., Watson C. I., Garris M. D. & Hicklin A. 2003 *Studies of Fingerprint Matching Using the NIST Verification Test Bed VTB* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology Internal Report 7020.

[3] Tabassi E., Wilson C. L. & Watson C. I. 2004 *Fingerprint Image Quality* (National Institute of Standards and Technology, Gaithersburg, MD), National Institute of Standards and Technology Internal Report 7151.

	Image Quality (i)							
	1	2	3	4	5	6	7	8
p^r	0.315	0.215	0.220	0.100	0.050	0.035	0.020	0.045
p^l	0.240	0.190	0.230	0.120	0.070	0.045	0.030	0.075
$p^s(\text{empirical})$	0.220	0.225	0.260	0.125	0.065	0.040	0.020	0.045
$p^s(\text{independence})$	0.479	0.253	0.183	0.052	0.018	0.008	0.004	0.003

Table 2: The probability mass functions for the image quality of left index fingers $p^l(i)$ (2 per person), right index fingers $p^r(i)$ (2 per person), the search quality $p^s(i)$, and the search quality if there was no environmental noise and independence between right and left index fingers.

	Quality (i)							
	1	2	3	4	5	6	7	8
d_i for $(t_1, t_2) = (1400, 2025)$	0.993	0.989	0.985	0.974	0.935	0.862	0.800	0.478
d_i for $(t_1, t_2) = (1300, 1880)$	0.995	0.992	0.989	0.980	0.945	0.885	0.825	0.521

Table 3: Detection probabilities d_i at the two threshold values for different qualities. Taken from figure 12 of ref. 1. The upper-right entry of the table was obtained by personal communication with NIST.

Quality	λ shape	θ scale	mean	std. dev.	median
1	2.8818	1278.1	3683.2	2169.7	3267.1
2	2.7774	1230.7	3418.1	2051.0	3017.8
3	2.7596	1155.7	3189.1	1919.8	2813.3
4	2.3225	1341.2	3114.9	2043.9	2681.0
5	1.4052	2307.3	3242.3	2735.1	2513.8
6	1.8624	1072.1	1996.7	1463.1	1652.9
7	1.3100	1474.4	1931.5	1687.5	1468.3
8	1.2187	834.2	1016.7	921.0	756.1

Table 4: Parameter values for the eight intraperson score distributions.

	Quality (i)							
	1	2	3	4	5	6	7	8
$f_i(\times 10^{-4})$ for $(\mathrm{t}_1, \mathrm{t}_2) = (1400, 2025)$	4.58	5.23	6.87	6.57	13.11	9.00	16.07	26.03
$f_i(\times 10^{-3})$ for $(\mathrm{t}_1, \mathrm{t}_2) = (1300, 1880)$	2.396	2.623	2.870	3.594	2.870	5.725	6.567	5.578

Table 5: False positive probabilities f_i at the two threshold values for different qualities. Taken from figure 12 of ref. 1. The upper-right entry of the table was obtained by personal communication with NIST.

Quality	μ scale	σ shape	mean
1	6.1716	0.1501	484.40
2	5.7506	0.2257	322.49
3	5.1832	0.3178	187.49
4	5.6222	0.2496	285.26
5	2.9135	0.6816	23.24
6	5.6804	0.2435	301.88
7	5.1847	0.3258	188.25
8	3.4073	0.6148	36.46

Table 6: Parameter values for the eight interperson score distributions.

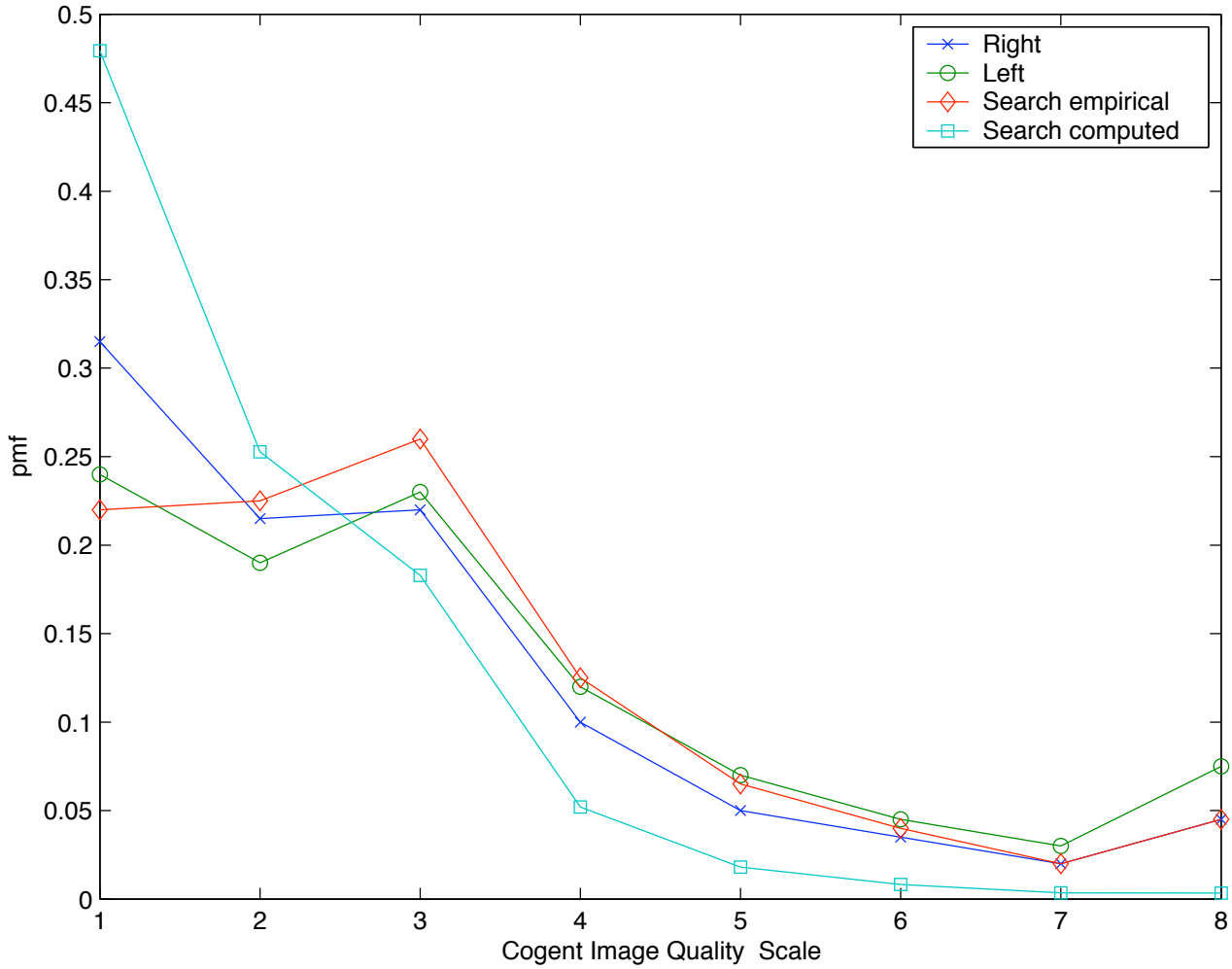

Figure 3: The probability mass functions for the image quality of left index fingers $p^l(i)$ (two per person), right index fingers $p^r(i)$ (two per person), the empirical search quality $p^s(i)$, and the computed search quality if there was independence between right and left index fingers and no operational noise. The first three curves are taken from figure 1 of ref. 1.

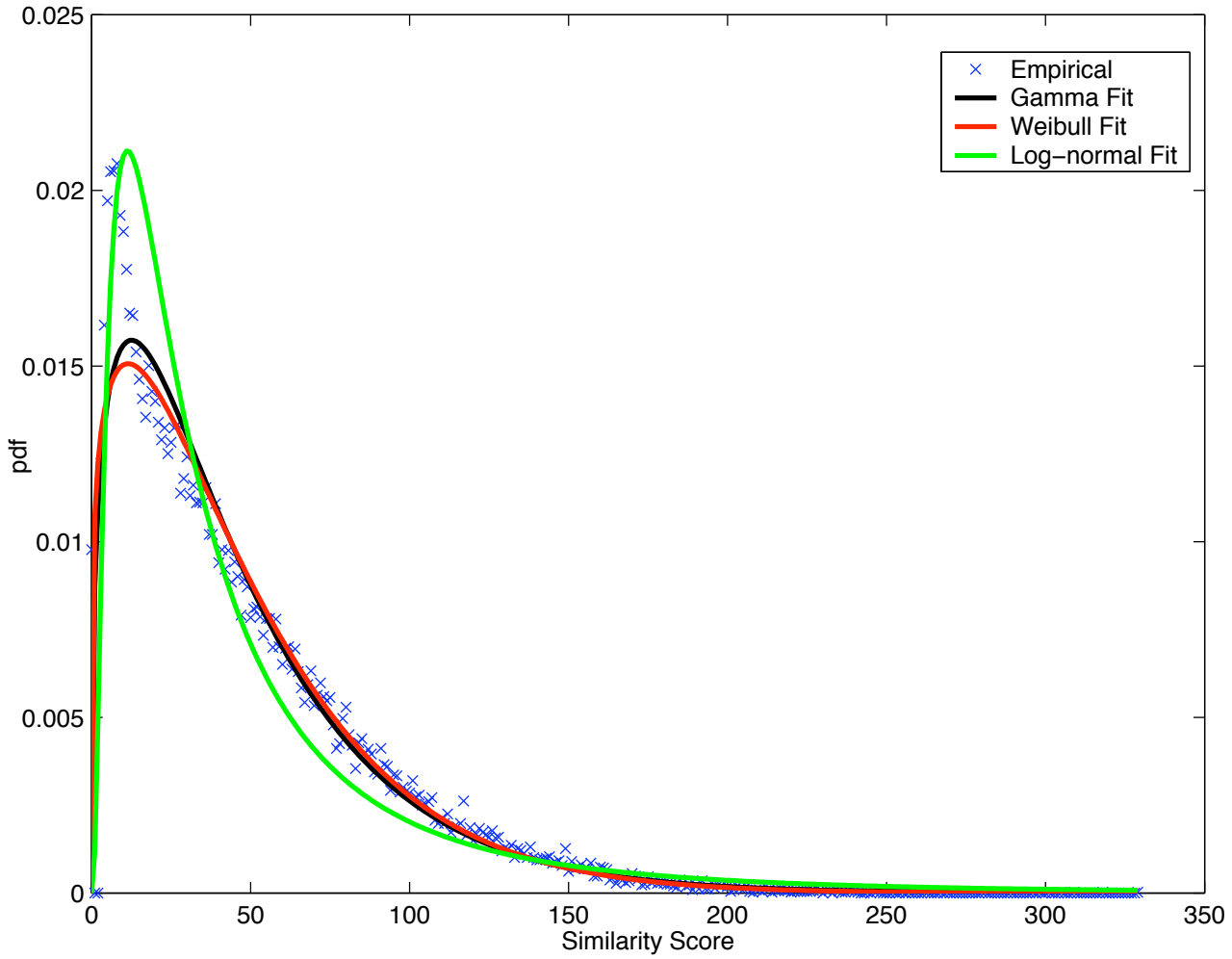

Figure 4: The pdf of the intraperson similarity scores from section 5.13 of ref. 2, along with the best-fit log-normal, Weibull and gamma distributions.

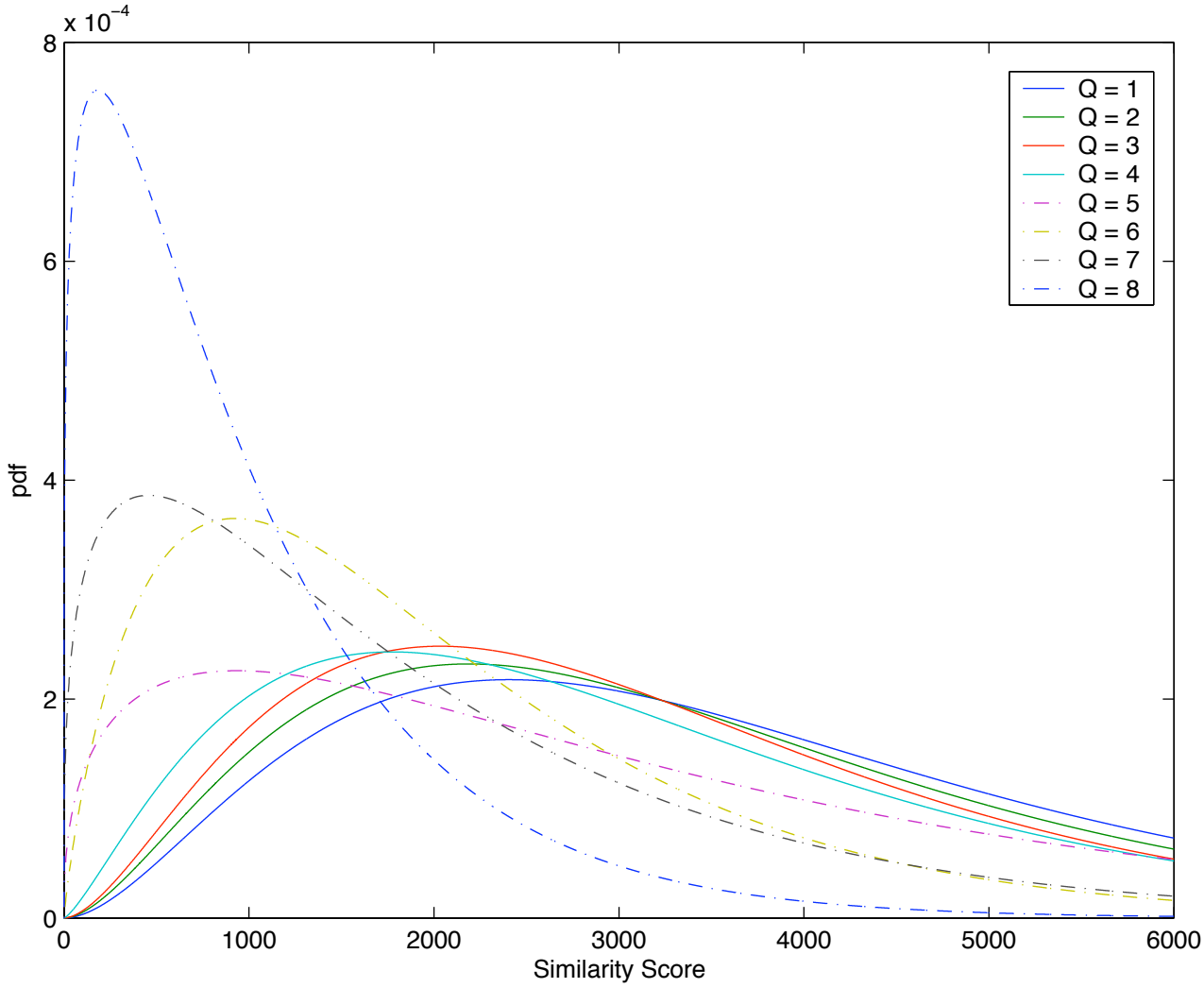

Figure 5: The eight intraperson similarity score distributions.

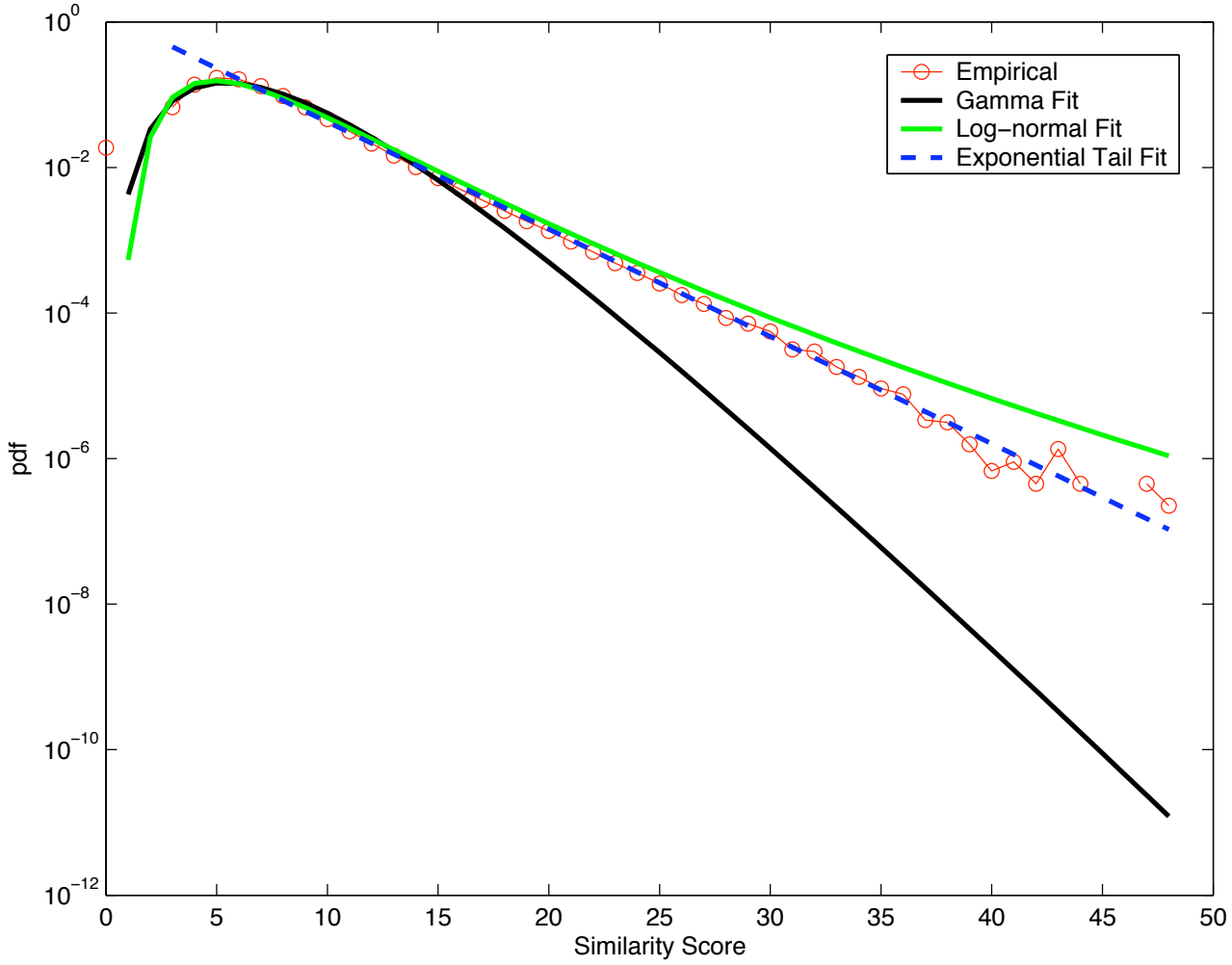

Figure 6: The pdf of the interperson similarity scores from section 5.13 of ref. 2, along with the best-fit log-normal distribution, gamma distribution and exponential tail.

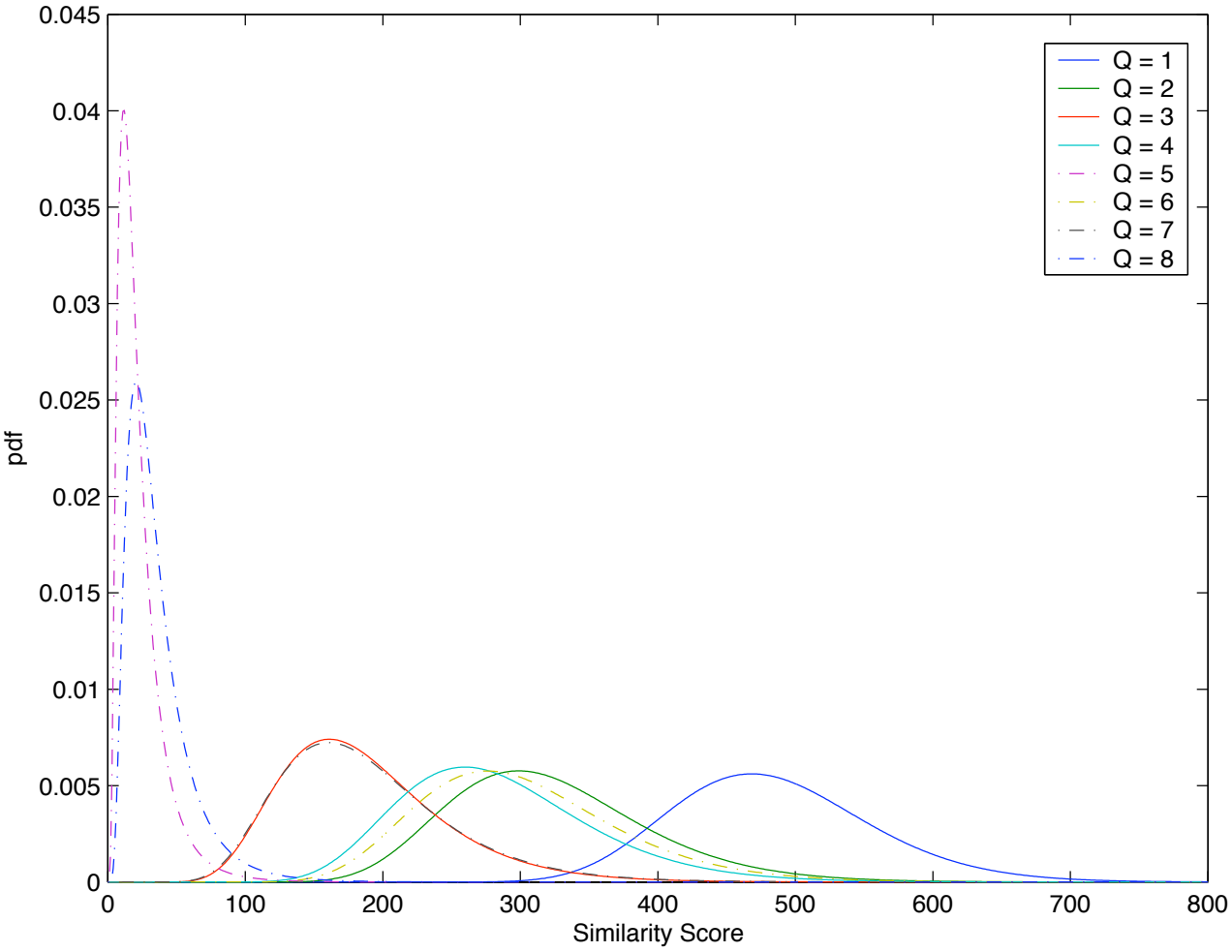

Figure 7: The eight interperson similarity score distributions.