The Myth of Goats:
How many people have fingerprints that are hard to match?

Austin Hicklin, Mitretek Systems
Craig Watson, NIST
Brad Ulery, Mitretek Systems

The Myth of Goats:
How many people have fingerprints that are hard to match?

Austin Hicklin
Brad Ulery
Mitretek Systems

Craig Watson
U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of
Standards and Technology
Information Access Division
Gaithersburg, MD 20899

September 2005

The Myth of Goats:
How many people have fingerprints that are hard to match?

Austin Hicklin, Mitretek Systems

Craig Watson, NIST

Brad Ulery, Mitretek Systems

13 September 2005

Abstract

The proportion of people who have fingerprints that are particularly hard to match (also known as "Goats") is a topic of great interest in biometrics, especially for those involved in the design, development, or evaluation of fingerprint-based identification or verification systems. There have been a variety of statements made in the recent past that a small percentage of people (usually 2%) cannot be fingerprinted due to poor quality fingers. This study shows these statements are based on misconceptions: the fact that some small percentage of fingerprints may be hard to match does not mean that a corresponding percentage of people are hard to match.

This study describes the results of tests using fingerprint data collected operationally by US-VISIT. Ten sets of right and left index fingerprints from each of 6,000 individuals were used in the evaluation. Two of the more accurate matchers from the NIST Software Development Kit (SDK) tests were used.

The definition of a Goat, or person whose fingerprints are intrinsically hard to match, varies. However, results clearly show that the proportion of Goats is very small, regardless of the definition. None of the 6,000 subjects had fingers that were <u>always</u> hard to match (with single-finger mate scores worse than a threshold corresponding to a verification False Accept Rate of 1%); less than 0.05% of the subjects had fingers that were <u>usually</u> hard to match; less than 0.3% of the subjects had fingers that were hard to match even a quarter of the time.

Many individuals were particularly easy to match: for 77% to 81% of subjects, <u>every</u> fingerprint comparison had mate scores better than a threshold corresponding to a verification False Accept Rate of 10^{-6} (0.0001%).

This study concludes that for the subject population (frequent users of US-VISIT) fingerprints that are hard to match <u>cannot</u> generally be attributed to intrinsic characteristics of a person's fingerprints, but should be attributed to collection problems or other characteristics of the specific fingerprints used. Note that these results should be generalized with caution: results obtained using less accurate matchers, data from a source with lower operational quality controls, or substantively different subject populations would be likely to differ.

Contents

Disclaimer

1 Introduction

Failures to match are obviously of serious concern in biometric systems. Every biometric system encounters some cases that cannot be acquired, enrolled, and/or matched. The reasons for these failures can generally be attributed to collection problems (due to problems with the operator, subject, collection devices, or software), or to intrinsic characteristics of the subject. This distinction is important, because failures due to collection problems can be lessened or eliminated with better quality control, but failures due to intrinsic characteristics of the subject limit the effectiveness of that biometric.

Various statements have been made about the proportion of the general population that is difficult to match using fingerprints: 2% is most frequently cited. For large-scale operational systems such as US-VISIT, 2% of the overall population is a very large number of people, and therefore of concern. This study attempts to verify or contradict the assumption that there is such a proportion of the population that is difficult to match using fingerprints.

The concept of a biometric menagerie was introduced in 1997 by Joseph Campbell[1] and expanded by George Doddington[2], defining

- sheep (subjects who are easily matched),
- goats (subjects who are particularly difficult to match),
- lambs (subjects exceptionally vulnerable to impersonation), and
- wolves (subjects exceptionally successful at impersonation).

The term "Goat" is now widely used in biometrics to refer to a person who is intrinsically difficult to match.

2 Background

There have been a variety of statements made in the recent past that a small percentage of people cannot be fingerprinted due to poor quality fingers. Most of these cite (directly or indirectly) a 2002 NIST report,[3] which made the following statements:

---

[1] See References [Campbell]
[2] See References [Doddington]
[3] See References [NIST2002]

(A) *"Tests by NIST using INS data show that for approximately 2% of the fingers in the INS database, the friction ridges are too damaged to be matched with existing technology." [pg. 2]*

(B) *"We conclude that approximately 2% of individuals in the general population may not be easily fingerprinted." [pg. 15]*

(C) *"Not all subjects can be easily fingerprinted with existing technology resulting in a 2% failure to acquire rate." [pg.21]*

A 2003 GAO Report,[4] citing the above NIST report, stated,

(D) *"Approximately 2 percent of the population cannot provide good fingerprint images." [pg 11]*

The NIST report was partly based on a 2000 Mitretek study for DOJ,[5] which stated,

(E) *"The quality of approximately 2 percent of INS IDENT flat fingerprints is so poor that it renders them virtually impossible to match using current IAFIS technology, and an additional 3 percent would be very unlikely to match." [pg vii]*

Each of these statements can be misinterpreted or misapplied. There are several issues involved here, which are key to defining what Goats are:

Problem fingerprints as opposed to people with problem fingers

The fact that a given percentage of fingerprints in a database may be hard to match does not mean that a corresponding percentage of people have fingers that are always hard to match. A fingerprint is often hard to match for a variety of reasons that are specific to that individual image, such as poor quality control in data collection, that do not reflect on the ability of the person to provide good fingerprints in the future. Every person can produce poor-quality fingerprints. Note that all of the statements cite the same 2% figure even though statements (A), (C) and (E) refer to fingerprints, and statements (B) and (D) refer to people.

Population and data source

All of these statements refer to INS[6] recidivist data, collected by what was then INS (now BICE[7]) from illegal immigrants, under

---

[4] See References [GAO]

[5] See References [IQS]

[6] Immigration and Naturalization Service, U.S. Department of Justice

[7] Bureau of Immigration and Customs Enforcement, U.S. Department of Homeland Security

difficult operational conditions, using fingerprint scanners with less-than-ideal image quality. Statistics based on this population (mostly composed of manual laborers) cannot be extrapolated to the overall human population, and the distribution of image qualities cannot be assumed to apply to other sources of data. Some population characteristics that are known or believed to be associated with lower fingerprint matching accuracy include manual labor that damages the skin (through abrasion or chemicals), age, and small physical size. Data sources that come from populations that differ substantially in these terms should be expected to differ in ease of matching.

Matching technology

None of the statements was based on state-of-the-art technology for the types of fingerprints used. The NIST test was based on the VTB,[8] which was shown to be roughly average in performance when compared against a range of other matchers in the FpVTE[9] and SDK[10] evaluations. The results in (D) used flat fingerprints on a system specifically optimized for rolled fingerprints. Some systems available today are much more accurate, especially with poor-quality single-finger flat images. Determining which individuals are Goats also depends on the matching algorithm in ways beyond overall accuracy: fingerprints that cause problems for any one algorithm will overlap with fingerprints that cause problems for others, but will often include some cases that are particularly difficult solely for that algorithm.

Difficulty of matching

Except in extreme cases, there is no absolute line defining fingerprints that are hard to match from those that are easy to match: hard (or easy) to match could be defined in a variety of ways. Goats could be defined in absolute terms, as people whose fingerprints are consistently matched below a given threshold[11], based on evaluation and operational experience. A different definition of Goats could be in relative terms, defined as people

---

[8] The NIST Verification Test Bed (VTB) is an export-controlled open source fingerprint matcher: see References [VTB]

[9] See References [FpVTE]

[10] See References [SDK]

[11] Matcher score thresholds are set to trade off false matches against false rejects.

who are harder to match than others, but who may or may not actually be hard to match in absolute terms.

Number of fingers

> For biometrics such as iris or fingerprint, for which one subject can provide multiple samples (two eyes or ten fingers), failures to match can be based on each sample, or each subject. Many people have hard to match little fingers, but very few have ten hard to match fingers: does a Goat have one poor-quality finger, or are all ten fingers poor quality?

Many evaluations have been conducted that can be used to determine the proportion of the fingerprints in the test that were difficult to match, using a specific set of matcher technologies. In most evaluations the proportion of fingerprints that failed to match (false rejects) can readily be determined, for a given matcher, at a given threshold or false accept rate (FAR). Unfortunately, without multiple fingerprints per finger (and per person), these studies reflect the proportion of problem fingerprints, and say nothing about whether these were caused by intrinsic characteristics of the subjects.

The only study known to have evaluated the consistency of intra-subject fingerprint matching accuracy is the small-scale study reported in [NIST2002]. In that study, 100 subjects with 30 sets of right and left index fingerprints each were evaluated using the VTB matcher. Two of the subjects were found to be problematic:

> *"Two of the match blocks [i.e. subjects] had average match scores that were so low that they were only marginally distinguishable from non-match blocks. When the 30 images of each of the marginal match blocks were examined, it was found that the friction ridges of these individuals were abraded to the point where no ridges were present. This did not appear to be an equipment problem. The images had sufficient contrast and were not blurred. The fingers did not have detectable or repeatable friction ridge patterns." [pg 14-15]*

The NIST study was appropriately designed and conducted, but as stated above there were several issues that limited how the results could be generally applied:

- Only 100 subjects were available. This sample size is not large enough to make a definitive statement about a 2% rate with high confidence.

This limits the level of certainty of measurements, so that 2/100 can only be stated as 2% ± 2.7% (95% confidence)[12], or 0.0% to 4.7%.

- The matcher used was not one of the more accurate matchers available.
- The population involved was limited to illegal aliens apprehended on the Mexican-US border, most of whom were manual laborers.
- The capture devices used were not representative of the more accurate livescan devices now available.
- Operational quality control for fingerprint capture was limited.

No studies are known to have been conducted to determine the proportion of people in the general population with hard to match fingerprints, using a large number of subjects with multiple fingerprints per finger, and accurate matchers. The current study seeks to fill this gap, and expand upon the previous NIST work.

3   Experiment Description

The primary objectives of this study were to determine the proportion of the US-VISIT population for whom

- One finger is always or usually particularly difficult to match
- Both fingers are always or usually particularly difficult to match
- One finger is always or usually particularly easy to match
- Both fingers are always or usually particularly easy to match

3.1   Data

NIST has a large variety of fingerprint data available. One set in particular, the US-VISIT Point of Entry dataset (POE) contains all fingerprints collected by US-VISIT between January and June 2004. The POE dataset included left and right index fingerprints from 4.2 million subjects. Note that no Failures to Enroll (FTEs) were excluded from this data: the overall dataset includes every single non-amputee who entered the United States with a U.S. Visa.[13] All mating information (knowing which fingerprints belonged to the same person) was determined through document numbers, not through use of a fingerprint matcher. Two subsets of the POE data were used in this study:

---

[12] This is assuming a binomial distribution; otherwise, the confidence bands will be larger.

[13] Amputees account for much less than 0.1% of the US-VISIT population. [ref. Neal Latta, US-VISIT]

Multi-Instance Subset of POE

Of the 4.2 million subjects, approximately 6,600 had ten or more sets of fingerprints in the POE dataset. From these, ten sets of fingerprints (left and right index) were selected from each of 6,000 subjects.[14] Two of the more accurate SDK matchers (Matchers H and I) were used to match each finger for a subject against all of the other corresponding fingers from the same subject. Since ten sets of fingerprints were available from each subject, each matcher generated 90 scores per finger, or 360 scores per person: the Multi-Instance Subset of POE included 2,160,000 scores in all.[15]

Note that this population represents individuals who were processed by US-VISIT ten or more times in a six month period. While the overall US-VISIT population is to some extent a cross-section of the world's population, this multi-instance population is primarily composed of flight crews and business travelers.

Overall Subset of POE

Another random selection of 61,996 subjects[16] from the overall US-VISIT-POE dataset was made, with two sets of fingerprints (left and right index) from each subject. The score distributions from this dataset from the H and I matchers were used to determine the extent to which the Multi-Instance Subset of POE corresponded to the overall US-VISIT population.

3.2   Thresholds

For this study, thresholds were required to correspond to the nebulous concepts of "Hard to match" and "Easy to match." The thresholds used were not absolute: whether something is hard to match is an application-dependent concept. As rough guidelines for the purposes of this study,

---

[14] Both subjects and sets of fingerprints were randomly selected. 6,000 subjects were used because 6,000-subject matrices were standard in the SDK tests.

[15] The matchers used were not fully symmetric: matching A against B generally results in scores that are similar to but not precisely the same as matching B against A. Of the 90 scores per finger, therefore, 45 were redundant.

[16] 62,000 subjects were randomly selected; 4 subjects could not be processed due to invalid file formatting.

- scores associated with a verification FAR[17] of 10^{-2} or worse can be considered hard to match, and
- scores associated with a verification FAR of 10^{-6} or better can be considered easy to match.[18]

The categories used in this study to define difficulty of matching in terms of FAR are shown in Table 1.

Score Thresholds in terms of False Accept Rates	Categories used in this study
FAR=1	Unusable
$10^{0} >$ FAR $\geq 10^{-2}$	Hard to match
$10^{-2} >$ FAR $\geq 10^{-6}$	Debatable (between hard and easy to match)
FAR $< 10^{-6}$	Easy to match

Table 1: Categories of thresholds, defined in terms of 1:1 False Accept Rates

For a subject/finger to be <u>always</u> hard to match, the maximum score of all of the cross-comparisons must be worse than a specific score threshold. For a subject/finger to be <u>usually</u> hard to match, the median score of all of the cross-comparisons must be worse than a specific score threshold.

The mate and non-mate score distributions were determined for the POE data using two of the most accurate matchers from the NIST SDK evaluations,[19] Matcher H and Matcher I. For reference, the associations between False Accept and True Accept Rates (as calculated in the SDK tests on the US-VISIT POE data) are shown in Table 2.

---

[17] The False Accept Rates stated throughout the paper are 1:1 (verification) FARs, based on single-finger matching.

[18] Note that a verification FAR of 10^{-6} on a single-finger search would not by itself translate to a definitive search against a large identification system (with a gallery in the millions). Such systems rely on the use of multiple fingers to achieve those levels of accuracy.

[19] See References: [SDK]

False and True Accept Rates					
FAR		**TAR**			
		Matcher I		**Matcher H**	
		R.Index	**L.Index**	**R.Index**	**L.Index**
1	10^0	100.0%	100.0%	100.0%	100.0%
0.1	10^{-1}	99.5%	99.5%	99.7%	99.7%
0.01	10^{-2}	99.2%	99.2%	99.4%	99.3%
0.001	10^{-3}	99.0%	98.9%	99.2%	99.1%
0.0001	10^{-4}	98.7%	98.5%	98.9%	98.7%
0.00001	10^{-5}	98.3%	98.0%	98.5%	98.0%
0.000001	10^{-6}	97.9%	97.2%	97.9%	97.4%

Table 2: False and True Accept Rates for the matchers used in the study on US-VISIT POE Data, as measured in the NIST SDK tests

4   Distribution of Mate Scores

This section describes the key results of the study, the overall distribution of mate scores and the distribution of mate scores by subject.

4.1   Overall Distribution of Mate Scores

Since the population in the multi-instance dataset differs to some extent from the overall US-VISIT population, it is necessary to demonstrate that the multi-instance and overall datasets correspond in their distributions, so that findings related to the multi-instance data can be reasonably expected to apply to the overall US-VISIT population.

Table 3 and Table 4 show the distribution of scores in terms of False Accept Rates for the overall US-VISIT dataset, and for the Multi-Instance Subset of POE. Note the degree of correspondence between the distributions.

Overall Subset of POE				
Distribution of scores in terms of False Accept Rates				
log(FAR)	Right Index		Left Index	
	Matcher H	Matcher I	Matcher H	Matcher I
0	0.60%	0.03%	0.48%	0.05%
0 to -1	0.03%	0.47%	0.05%	0.36%
-1 to -2	0.04%	0.22%	0.07%	0.34%
-2 to -3	0.20%	0.23%	0.36%	0.41%
-3 to -4	0.28%	0.28%	0.41%	0.46%
-4 to -5	0.30%	0.38%	0.52%	0.68%
-5 to -6	0.41%	0.59%	0.69%	0.96%
<-6	98.16%	97.81%	97.41%	96.74%
Scores	61,996	61,996	61,996	61,996
Subjects	61,996	61,996	61,996	61,996

Table 3: Distribution of Overall Subset of POE scores in terms of False Accept Rates

Multi-Instance Subset of POE				
Distribution of scores in terms of False Accept Rates				
log(FAR)	Right Index		Left Index	
	Matcher H	Matcher I	Matcher H	Matcher I
0	0.56%	0.03%	0.40%	0.04%
0 to -1	0.04%	0.46%	0.05%	0.30%
-1 to -2	0.05%	0.24%	0.06%	0.28%
-2 to -3	0.22%	0.24%	0.27%	0.30%
-3 to -4	0.30%	0.30%	0.34%	0.38%
-4 to -5	0.35%	0.43%	0.41%	0.52%
-5 to -6	0.48%	0.63%	0.56%	0.80%
<-6	98.01%	97.66%	97.92%	97.37%
Scores	540,000	540,000	540,000	540,000
Subjects	6,000	6,000	6,000	6,000

Table 4: Distribution of Multi-Instance Subset of POE scores in terms of False Accept Rates

These results are shown graphically in Figure 1.
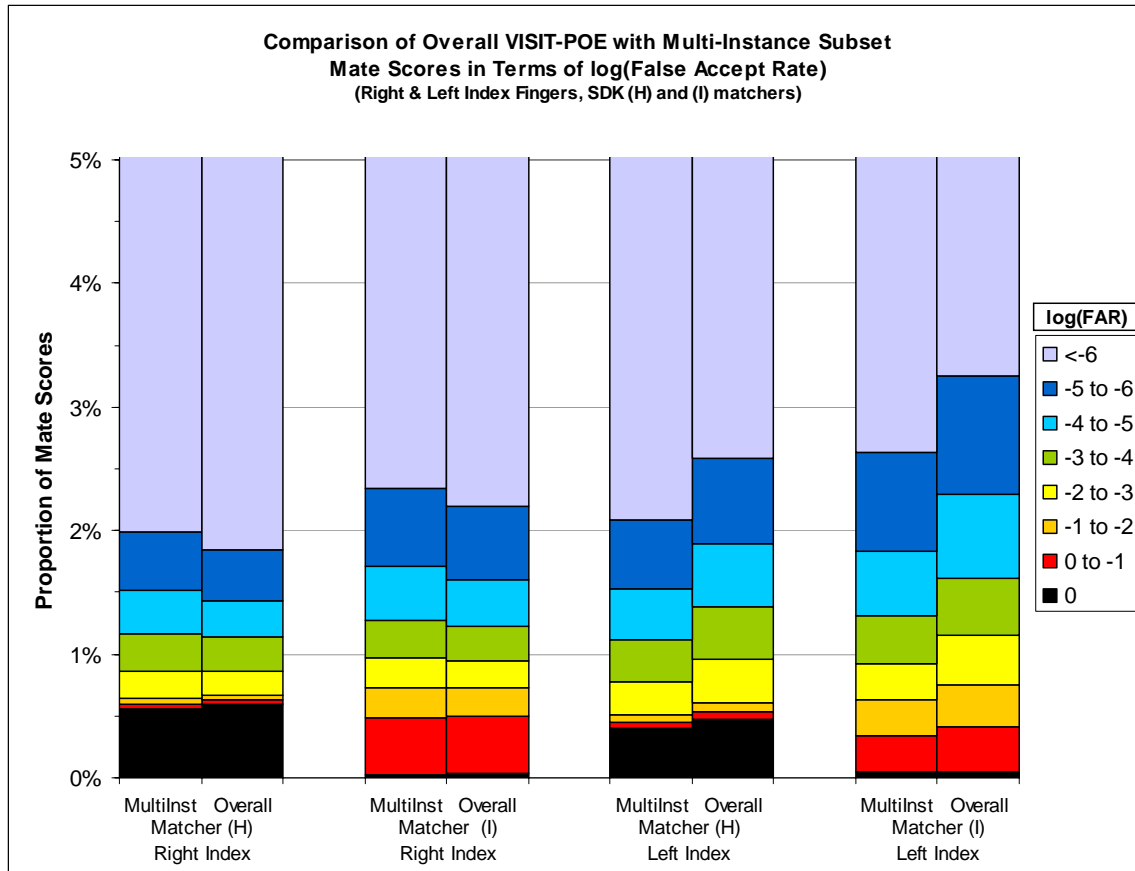
Figure 1: Comparison of matcher scores from the Overall and Multi-Instance subsets of US-VISIT-POE data

Further analysis of the actual scores in the tails of the distributions (e.g. the 1st percentile and 0.1 percentile) also showed close correspondence.

4.2 Distribution of Mate Scores by Subject

When the 10 fingerprints for each person's finger were matched, all cross-comparisons yielded 90 scores for each matcher. This analysis is based on the distribution of the minimum, 1st quartile, median, average, 3rd quartile, and maximum values of these scores.[20] Figure 2 shows the median score, interquartile distance (difference between 1st and 3rd quartiles), minima, and maxima for all right index fingers, sorted by median, using the Matcher H matcher.[21] Note that the scores vary substantially, both in terms of median values and inter-quartile distance, but that only a very small proportion of the scores fall within the FAR range of 1 to 10^{-6}: the

---

[20] The redundancy in the scores (inclusion of both A:B and B:A comparisons) does not affect these measures, but would affect measures such as standard deviation.

[21] The actual matcher scores are omitted, as they are considered proprietary.

overwhelming majority of scores are far better than even the easy to match threshold. Note in Figure 2 that the threshold of FAR=1 cannot be differentiated from FAR=10^{-2} at this scale.
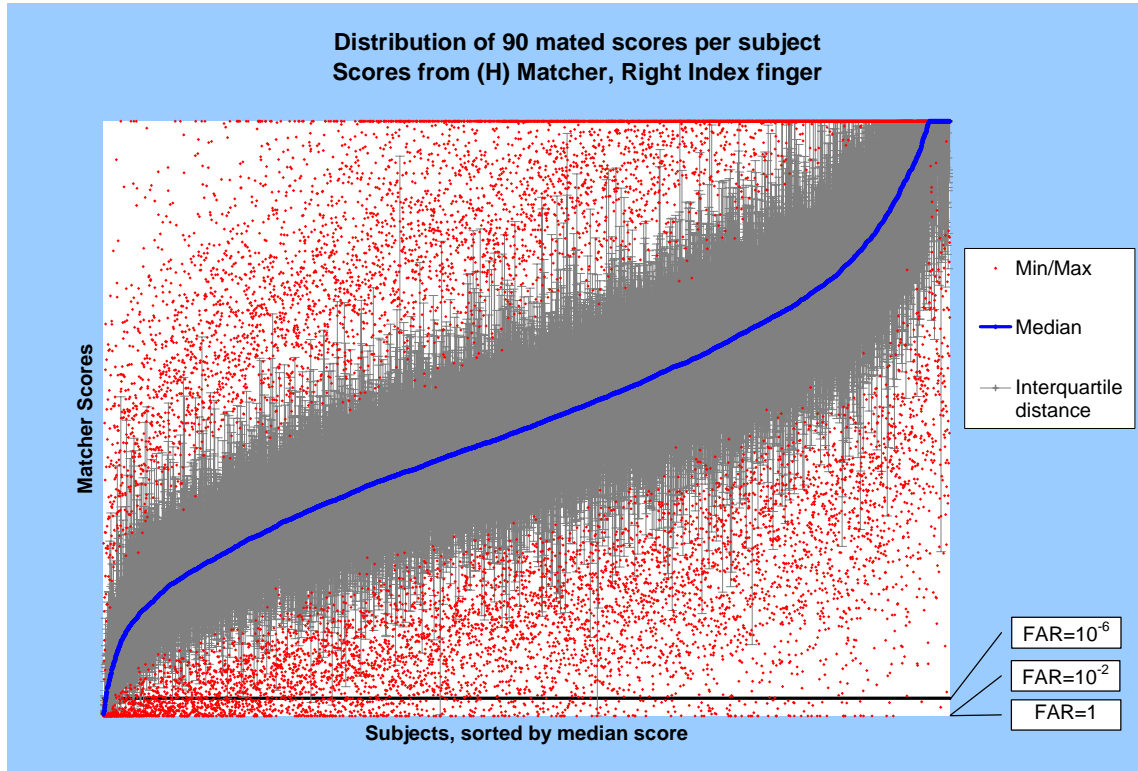
Figure 2: Distribution of matcher scores from the (H) matcher (right index fingers), showing median scores (blue line), 1st and 3rd quartiles (gray lines), and minima/maxima (red points)

There is a clear relationship between subject and the distribution of matcher scores: in other words, some people are harder (or easier) to match than others.

Table 5 and Table 6 show greater detail for the score distributions. The highlighted regions of these tables show that:

No subject was always hard to match

> No subjects (for either finger, either matcher) had maximum scores worse than a FAR of 10^{-2}.

Less than 0.03% of subjects had fingers that were usually hard to match

For the H matcher, no subject (for either finger) had a median score worse than a FAR of 10^{-2}, but for the I matcher, 0.02%-0.03% of the population[22] had median scores worse than a FAR of 10^{-2}.

Less than 0.3% of subjects had fingers that were hard to match even a quarter of the time.

Between 0.08% and 0.32% of the population had a 1st Quartile worse than a FAR of 10^{-2}.

6.3% of subjects had fingers that were ever hard to match.

The minimum score for 6% of subjects was worse than a FAR of 10^{-2}. Most hard to match fingerprints are the worst scores for their subjects, and are not generally representative of their subjects. Note that the H and I matcher differed substantially on the proportion of unusable (FAR=1) scores.

About 80% of subjects were always easy to match

- 77.0-80.9% of the subjects had a minimum score better than a threshold of 10^{-6}.

More than 99.5% of subjects were usually easy to match

- 99.5-99.7% of the subjects had a median score better than a threshold of 10^{-6}.

---

[22] Ranges show range of measured values for the two matchers and left and right index fingers.

| Finger | Matcher | Type | Summary of Cumulative Distribution, Right Index | | | | | | | Easy to match |
| | | | Hard to match | | | | | | | |
			FAR=1	FAR≥10^{-1}	FAR≥10^{-2}	FAR≥10^{-3}	FAR≥10^{-4}	FAR≥10^{-5}	FAR≥10^{-6}	FAR<10^{-6}
2 (Right Index)	Matcher H	Min	5.15%	5.68%	6.33%	8.18%	11.10%	14.32%	19.07%	80.93%
		Quartile1	0.02%	0.05%	0.08%	0.32%	0.65%	0.95%	1.38%	98.62%
		Median	-	-	-	0.02%	0.13%	0.18%	0.30%	99.70%
		Average	-	-	-	-	0.03%	0.07%	0.13%	99.87%
		Quartile3	-	-	-	-	0.03%	0.08%	0.12%	99.88%
		Max	-	-	-	-	-	0.02%	0.03%	99.97%
	Matcher I	Min	0.48%	4.28%	6.33%	8.68%	11.87%	15.95%	21.05%	78.95%
		Quartile1	0.02%	0.05%	0.22%	0.47%	0.68%	1.23%	1.80%	98.20%
		Median	-	-	0.03%	0.10%	0.17%	0.28%	0.43%	99.57%
		Average	-	-	0.03%	0.07%	0.13%	0.22%	0.37%	99.63%
		Quartile3	-	-	-	-	0.08%	0.10%	0.17%	99.83%
		Max	-	-	-	-	-	-	0.02%	99.98%

Table 5: Summary of Distribution of matcher scores, right index fingers

| Finger | Matcher | Type | Summary of Cumulative Distribution, Left Index | | | | | | | Easy to match |
| | | | Hard to match | | | | | | | |
			FAR=1	FAR≥10^{-1}	FAR≥10^{-2}	FAR≥10^{-3}	FAR≥10^{-4}	FAR≥10^{-5}	FAR≥10^{-6}	FAR<10^{-6}
7 (Left Index)	Matcher H	Min	5.05%	5.68%	6.40%	8.98%	11.97%	15.80%	20.17%	79.83%
		Quartile1	0.05%	0.12%	0.15%	0.33%	0.58%	0.88%	1.42%	98.58%
		Median	-	-	-	0.05%	0.17%	0.23%	0.27%	99.73%
		Average	-	-	-	-	-	0.05%	0.10%	99.90%
		Quartile3	-	-	-	-	0.02%	0.07%	0.08%	99.92%
		Max	-	-	-	-	-	-	-	100.00%
	Matcher I	Min	0.72%	4.00%	6.28%	9.30%	12.77%	17.28%	23.03%	76.97%
		Quartile1	0.02%	0.10%	0.32%	0.57%	0.82%	1.38%	2.45%	97.55%
		Median	-	-	0.02%	0.13%	0.22%	0.38%	0.55%	99.45%
		Average	-	-	-	0.05%	0.10%	0.27%	0.52%	99.48%
		Quartile3	-	-	-	0.02%	0.05%	0.08%	0.13%	99.87%
		Max	-	-	-	-	-	-	-	100.00%

Table 6: Summary of Distribution of matcher scores, left index fingers

The confidence intervals for the summary numbers highlighted above are detailed in Table 7.[23] The highest and lowest observed numbers of subjects among the two matchers and two hands are included.

---

[23] Confidence intervals for unobserved events were calculated using the intervals for a single instance as a bound.

	Number of subjects		Confidence Intervals (95%)							
	Low	High	Low			High			From	To
Always hard to match	0	0	0.02%	±	0.03%	0.02%	±	0.03%	0.00%	0.05%
Usually hard to match	0	2	0.02%	±	0.03%	0.03%	±	0.05%	0.00%	0.08%
Hard to match 1/4 of the time	4	18	0.07%	±	0.07%	0.30%	±	0.14%	0.00%	0.44%
Always easy to match	4618	4856	76.97%	±	1.07%	80.93%	±	0.99%	75.90%	81.93%
Usually easy to match	5967	5984	99.45%	±	0.19%	99.73%	±	0.13%	99.26%	99.86%
Easy to match 3/4 of the time	5990	5995	99.83%	±	0.10%	99.92%	±	0.07%	99.73%	99.99%
	(out of 6,000 subjects, two matchers, right and left index fingers)									

Table 7: Summary of results with confidence intervals

When Doddington introduced the concept of Goats, he defined them "as those speakers who are particularly difficult to recognize." [pg 1] He then determined the existence of Goats using analyses of variance (F-test and Kruskal-Wallis Test). The subtle distinction here is that he defined Goats in absolute terms (difficult to recognize), but conducted the test in relative terms (determining if some of the subjects were more difficult than others to recognize). In the field of speaker recognition, this is probably not a meaningful distinction. However, with fingerprints, the accuracy level is high enough that some subjects can be more difficult to recognize than others (in relative terms), but still easy to recognize (in absolute terms). There are certainly individuals whose fingers are harder to match than others. This can be shown through analyses of variance, or more simply through the results in Figure 2, which shows that the median score and range of scores are clearly dependent on subject.

5   Image Quality Values

This study defines Goats as individuals who are difficult to match in terms of matcher scores. A different study could define Goats as individuals with poor-quality fingerprints, in terms of image quality metrics (IQMs) rather than through matcher scores. IQMs are practical to use operationally, whereas matcher scores are generally only available for after-the-fact evaluation.

Table 8 shows the overall distribution of the NIST Fingerprint Image Quality (NFIQ) values[24] for the Multi-Instance Subset of POE. NFIQ defines values of 4 or 5 as poor or very poor quality.

NFIQ	Right Index	Left Index
1	28.6%	33.8%
2	42.3%	37.8%
3	26.3%	25.4%
4	1.7%	1.6%
5	1.1%	1.4%

Table 8: Distribution of image quality as measured using NFIQ

The poor image quality values (NFIQ 4/5) were not distributed evenly across subjects. Table 9 shows the percentage of subjects having fingerprints with poor image quality values (for the 6,000 individuals with ten samples of fingerprints from each finger). For example, for 0.1% of the subjects, all ten of their right index fingerprints were poor quality.

# of NFIQ4/5		Right Index		Left Index	
None	0	85.6%	85.6%	84.9%	84.9%
Some	1	8.6%		9.1%	
	2	2.8%	13.3%	2.9%	13.9%
	3	1.2%		1.2%	
	4	0.8%		0.8%	
Most	5	0.5%		0.5%	
	6	0.2%		0.2%	
	7	0.2%	1.1%	0.3%	1.2%
	8	0.1%		0.2%	
	9	0.1%		0.1%	
All	10	0.1%	0.1%	0.1%	0.1%

Table 9: Prevalence of multiple poor-quality fingerprints

These results are far from random: if NFIQ values from Table 8 were distributed with no relation to subject, about 74% would have no NFIQ4/5 fingerprints[25], only 0.001% would have five out of ten NFIQ4/5 fingerprints, and only $6*10^{-16}$ would have ten out of ten NFIQ4/5 fingerprints. The results show that that image quality as measured by

[24] NFIQ is a standard, publicly-available method of measuring fingerprint image quality, described at ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf ; the software (export-controlled) is available at http://fingerprint.nist.gov/NFIS/.
[25] 74%: for each finger (averaging left and right), 97.1% of fingerprints are NFIQ<4; 0 out of 10 fingerprints with NFIQ4/5 = $0.971^{10} \approx 74\%$

NFIQ is not randomly distributed among subjects. Instead, NFIQ4/5 fingerprints are concentrated in a small group of subjects: a majority of the NFIQ4/5 fingerprints occur in 3.3% of subjects.

Note that the matchers used are capable of matching even poor quality fingerprints at relatively high scores. This explains the differences between the NFIQ results and the matcher score results. Less accurate matchers might show a greater degree of correspondence to the NFIQ results, with the result that a study using such matchers would measure a larger proportion of goats in the population.

6   Correspondence between Fingers

For subjects with poor-quality fingers, it is important to know whether poor-quality fingers on one hand are correlated with poor-quality fingers on the other hand.

Table 10 shows the number of poor-quality fingerprints per subject, as determined by NFIQ, counting both left and right index fingers. In other words, 2% of the 6,000 subjects had from 6 to 10 poor-quality fingerprints, out of 20 (10 left and 10 right index fingers), and for 0.6% of subjects a majority of fingerprints from both hands were poor quality.

Multi-instance POE subset	
# of fingerprints per subject with NFIQ≥4 (out of 10 left and 10 right index fingers)	
0	78.0%
1-5	19.5%
6-10	2.0%
11-15	0.5%
16-20	0.1%

Table 10: Number of poor-quality fingerprints per subject (10 left and 10 right index fingers)

Table 11 shows how matcher scores (in terms of FAR) correspond for fingerprints collected at the same time. For example, only 0.07%-0.08% of all sets of right and left index fingerprints were worse than 10^{-2} for both fingerprints. 99.5%-99.7% of subjects had at least one finger that was easy to match.

Note that conditional probabilities that can be drawn from these results:

- The overall probability of a hard to match left index finger is about 0.7% (0.65% for Matcher H, or 0.73% for Matcher I).

- If the right index finger is hard to match, the probability the left index finger collected at the same time is hard to match jumps to 13% (12.9%[26] for Matcher H, and 13.1% for Matcher I).

Summary of corresponding FAR distributions for fingerprints collected at one time				
Matcher H				
		Left Index		
		$> 10^{-2}$	10^{-3} to 10^{-6}	$< 10^{-6}$
Right index	$> 10^{-2}$	0.07%	0.04%	0.40%
	10^{-3} to 10^{-6}	0.05%	0.16%	1.38%
	$< 10^{-6}$	0.54%	1.14%	96.24%
Matcher I				
		Left Index		
		$> 10^{-2}$	10^{-3} to 10^{-6}	$< 10^{-6}$
Right index	$> 10^{-2}$	0.08%	0.08%	0.46%
	10^{-3} to 10^{-6}	0.07%	0.22%	1.72%
	$< 10^{-6}$	0.58%	1.32%	95.47%

Table 11: Correspondence between fingers

7   Correspondence between Matchers

The matchers used had similar levels of matching accuracy, but did not always return equivalent results. Table 12 shows the proportions of individual scores that were outside the hard/easy to match thresholds for each matcher separately and together.

Distribution of Individual Scores		Matcher H	Matcher I	Both
Hard to match	FAR $>=10^{-2}$	0.6%	0.7%	0.4%
Easy to match	FAR $<10^{-6}$	98.0%	97.5%	97.0%

Table 12: Correspondence between matchers

Table 13 shows the distribution of the differences in FAR between the two matchers, over all individual scores. For example, the 6 row means that the log of the FAR differed by 6, such as would occur if one matcher had a FAR of 10^{-7} and the other had a FAR of 10^{-1}. Note that there were cases in which one matcher returned a FAR of 1 (10^0) when the other returned a FAR of 10^{-7}.

---

[26] 12.9% = 0.07% / (0.07% + 0.04% + 0.40%)

Distribution of Individual Scores	
Difference between log(FAR$_H$) and log(FAR$_I$)	
0	97.26%
1	1.41%
2	0.73%
3	0.35%
4	0.16%
5	0.06%
6	0.02%
7	0.01%

Table 13: Difference between log(False accept rates) for matchers H and I

8   Conclusions

1   The results clearly show that the proportion of individuals who had consistently hard to match fingerprints is smaller than 0.05% in the US-VISIT population.

The study used a definition of "hard to match" as a matcher score corresponding to a False Accept Rate of 10^{-2} (1%) or worse, and a definition of "easy to match" as a matcher score corresponding to a False Accept Rate of 10^{-6} (0.0001%) or better.

- None of the 6,000 subjects had fingers that were always hard to match
- Less than 0.05% of the subjects had fingers that were hard to match the majority of the time
- No more than 0.2% of the subjects had fingers that were hard to match even a quarter of the time

2   Most individuals were consistently easy to match.

Most of the US-VISIT population can be readily matched even at a high score threshold.

- For 77% to 81% of subjects, every fingerprint comparison had matcher scores better than a False Accept Rate of 10^{-6} (0.0001%).
- For more than 98% of the subjects, three-fourths of fingerprint comparisons were easy to match.
- For more than 99.5% of the subjects, the majority of the fingerprint comparisons were easy to match.
- Every one of the 6,000 subjects had at least one fingerprint comparison that was easy to match.

3   Fingerprints that are hard to match cannot generally be attributed to intrinsic characteristics of a person's fingers.

Each finger used in the study had a range of performance: none was consistently difficult to match, and more than 99.9% were easy to match in some cases. A hard to match fingerprint, therefore, is indicative of problems with that specific fingerprint image, and does not mean another fingerprint from the same finger would be hard to match.

4   Poor quality fingerprints on one hand are correlated with poor quality fingerprints collected at the same time from the other hand.

Only 0.7% of subjects had hard to match left index fingers. However, for subjects with hard to match right fingers, 13% of subjects had hard to match left index fingers.

Note: These results should be applied to systems outside US-VISIT with caution.

The population in US-VISIT is to some extent a cross-section of the world's population, and therefore it is tempting to state that these results are generally applicable. The following cautions should be noted:

- The proportion of hard to match cases is correlated to the accuracy of the matchers used. Less accurate matchers cannot be expected to have the same results.
- US-VISIT has well-defined quality control procedures for fingerprint collection. Datasets with less rigorous quality control procedures, or collected on less accurate capture devices, will not yield the same results.
- Different subject populations will not yield the same results. In particular, populations that have a greater proportion of manual laborers or elderly can be expected to have a greater proportion of hard to match fingerprints.

9   References

Campbell	Campbell; *Speaker Recognition: A Tutorial*; Proceedings of the IEEE, Vol. 85, No. 9, September 1997. http://www.ee.columbia.edu/~patricia/papers/tutorials/tutorial.pdf
Doddington	Doddington, et al; *Sheep, Goats, Lambs And Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation*; http://www.nist.gov/speech/publications/papersrc/icslp_98.pdf
FpVTE	C. Wilson, et al; Fingerprint Vendor Technology Evaluation 2003, NISTIR 7123. June 2004. http://fpvte.nist.gov/
GAO	GAO; "Border Security: Challenges in Implementing Border Technology," Statement of Nancy Kingsbury, Managing Director Applied Research and Methods, Testimony Before the Subcommittee on Terrorism, Technology, and Homeland Security and Subcommittee on Border Security, Immigration, and Citizenship, Committee on the Judiciary, United States Senate; March 12, 2003. http://www.gao.gov/new.items/d03546t.pdf
IQS	R.A. Hicklin, et al; *Implications of the IDENT/IAFIS Image Quality Study for Visa Fingerprint Processing*, October 2002. http://www.mitretek.org/publications/biometrics/NIST-IQS.pdf
NFIQ	E. Tabassi, et al; Fingerprint Image Quality NIST IR 7151, August 2004. ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf
NFIS	C. Watson, et al; NIST Fingerprint Image Software. http://www.itl.nist.gov/iad/894.03/databases/defs/nist_nfis.html
NIST2002	NIST; *Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability*, November 13 2002. ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf
SDK	C. Watson, et al; NIST Fingerprint SDK (Software Development Kit) Testing, NISTIR 7119. June 2004. http://fingerprint.nist.gov/SDK/