Soft biometrics systems: Reliability and asymptotic bounds

Antitza Dantcheva, Jean-Luc Dugelay and Petros Elia

***Abstract*—This work presents a preliminary statistical analysis on the reliability of soft biometrics systems which employ multiple traits for human identification. The analysis places emphasis on the setting where identification errors occur mainly due to cross-subject interference, i.e., due to the event that subjects share similar facial and body characteristics. Finally asymptotic analysis provides bounds which insightfully interpret this statistical behavior.**

I. Introduction

Soft biometrics, as defined by Jain [1], [2], are those characteristics which provide weak biometrical information about an individual, but lack distinctiveness and permanence, and thus do not suffice to fully identify a person. This lack of distinctiveness can be partially overcome by employing multiple traits to classify individuals in pre-defined categories. This approach, which draws from Bertillon [3], is currently being embraced by the research community, with interesting work (cf. [6]–[10]) providing new aspects and methods.

The associated advantages of multi-trait soft biometrics systems (SBSs), over classical biometric systems, include:

- the ability to identify unknown subjects, based on descriptions given by humans,
- the ability to handle reduced sensor accuracy and furthermore acquire data in a non-obtrusive manner,
- the ability to process data in a computationally efficient manner.

At the same time though, the above advantages often come with different practical restrictions, such as on the number of detectable traits. This specific restriction in turn results in increased inter-subject interference, which will play an important role. We define interference as the random event where, within a randomly chosen set of subjects, the specific individual picked for authentication is indistinguishable from another subject in the same set, by sharing similar characteristics. It then becomes apparent that a measure of performance must go beyond the detectors' equal-error-rate measures used in classical biometrics, and should also account for the detrimental effect of interference. Building on the work in [5], the current work seeks to provide some mathematical analysis of reliability of general SBS, as well as to concisely bound the asymptotic behavior of pertinent statistical parameters that are identified to directly affect performance. This quantification seeks to provide a meaningful measure of the adequacy of a given SBS for real life applications.

This work was supported in part by the European Commission under contract FP7-215372 ACTIBIO.

A. Dantcheva, J.-L. Dugelay are with the Multimedia Communications Department, EURECOM, Sophia Antipolis, France (email: {dantchev, dugelay}@eurecom.fr) (tel: +33 49300 8144)

P. Elia is with the Mobile Communications Department, EURECOM, Sophia Antipolis, France (email: elia@eurecom.fr) (tel: +33 49300 8132)

A. Soft biometrics systems and operational scenario

This work examines the scenario where an SBS performs the task of identification of a person drawn uniformly and randomly from a randomly chosen set of N candidate subjects. We denote this randomly chosen N-tuple of people as $\mathbf{v}$, drawn from a sufficiently large population, and we denote by $\mathbf{v}(i)$, $i = 1, ..., N$ the i-th candidate belonging to the specific group $\mathbf{v}$.

An employed SBS is associated to the following variables: the number λ of soft biometric traits, the number μ_i of trait instances that each trait (labeled by) i can assume, and the *overall number of categories* $\rho = \Pi_{i=1}^{\lambda}\mu_i$. We also let $\Phi = [1, 2, \cdots, \rho]$ denote the indexed set of all feature combinations, i.e., the set of all categories that the SBS can identify.

Example *1:*

- Examples of sufficiently large populations include:
  - male population of Nice,
  - residents of Berlin.
- Examples of randomly chosen N-tuples $\mathbf{v}$ include:
  - the set of people who logged in to a specific public computer in Nice, two days ago,
  - the set of people whose pictures were captured by a video surveillance camera in Berlin, yesterday.
- Example SBS can identify $\lambda = 4$ different traits which include hair and eye colors as well as the presence of beard and glasses. The SBS can identify $\mu_1 = 5$ different hair colors, $\mu_2 = 5$ different eye colors, and of course we have that $\mu_3 = \mu_4 = 2$. As a result the SBS is endowed with the ability to identify $\rho = 100$ categories.
- Example categories, members of Φ, may include:
  - 'brown hair, blue eyes, no beard, no glasses' $\in \Phi$
  - 'black hair, blue eyes, no beard, glasses present' $\in \Phi$.

B. Reliability of soft biometrics systems

In the aforementioned operational setting of interest, the *reliability* of an SBS captures the probability of false identification of a randomly chosen person out of a random set of N subjects. In such a setting, the reliability of an SBS is generally related to the number of categories that the system can identify. Furthermore the performance of such systems can be a function of the degree with which these features/categories represent the chosen set (of subjects) over which identification will take place, as well as a function of the robustness with which these categories can be detected. Finally reliability is related to N, where a higher N corresponds to identifying a person among an increasingly large set of possibly similar-looking people.

C. Results

Section II introduces the operational setting of the SBS. In this setting, the *number of effective categories*, to be henceforth denoted by F, is identified as an important parameter related to *subject interference*, and is shown to directly affect the overall performance of an SBS. In Section III, Lemma 1 and Corollary 2 provide closed form expressions of the exact probability of system error for a given $\mathbf{v}$. The expressions reveal the somewhat surprising fact that, in the interference limited setting of high-sensor resolution (negligible number of detection errors), the reliability of an SBS is entirely defined by $F(\mathbf{v})$, and not by the distribution of categories characterizing the subjects in $\mathbf{v}$.

Lemma 3 then describes the probability of error averaged over N-tuples of subjects drawn from large populations. Section IV establishes, under a uniformity assumption, the statistical distribution and mean of F (Lemma 4 and Lemma 5) and the closed form expression of the probability of error, for the interference limited setting, averaged over all possible N-tuples of subjects (Theorem 6).

Finally towards establishing the scaling laws in soft biometrics systems in this specific operational setting, Section V, Lemma 8 succinctly bounds the statistical behavior of F over large populations. These bounds address the following practical question: if more funds are spent towards increasing the quality of an SBS by increasing ρ, then what reliability gains do we expect to see? Specifically, towards answering this, the work provides bounds on the probability of different interference patterns in soft biometrics systems. The bounds suggest that, under the interference limited assumption, doubling ρ will result in a doubly exponential reduction in the probability that a specific degree of interference will occur.

Section V-A provides intuition on the above bounds, Section VI provides some conclusions, and the Appendix provides the proofs.

II. Error events, interference, and effective categories

Let the randomly chosen subject for identification, belong in category $\phi \in \Phi$, i.e., the subject has the set of facial and body features that constitute category (labeled by) ϕ. The SBS first produces an estimate $\widehat{\phi}$ of ϕ, and based on this estimate, tries to identify the chosen subject, i.e., tries to establish which candidate in $\mathbf{v}$ corresponds to the chosen subject. An error occurs when the SBS fails to correctly identify the chosen subject, confusing him or her with another candidate from the current N-tuple $\mathbf{v}$. An error can hence certainly occur when the category is incorrectly estimated, i.e., when $\widehat{\phi} \neq \phi$, or can possibly occur when more than one candidate belongs in the same category as the chosen subject, i.e., when the chosen subject is essentially indistinguishable to the SBS from some other candidates in $\mathbf{v}$. We recall that subject $\mathbf{v}(i)$ interferes with subject $\mathbf{v}(j)$ whenever the two subjects belong in the same category.

For a given $\mathbf{v}$, let $S_\phi \subset \mathbf{v}$ be the set of subjects in $\mathbf{v}$ that belong in a specific category ϕ. Furthermore let S_0 denote the set of people in $\mathbf{v}$ that do not belong in any of the categories in Φ. We here note that no subject can simultaneously belong to two or more categories, but also note that it is entirely possible that $|S_\phi| = 0$, for some $\phi \in \Phi$. Hence an error is caused due to estimation noise (resulting in $\widehat{\phi} \neq \phi$), due to interference, or when the chosen candidate belongs in S_0.

For a given $\mathbf{v}$, let

$$F(\mathbf{v}) := |\{\phi \in \Phi : |S_\phi| > 0\}|$$

denote the number of effective categories, i.e., the number of (non-empty) categories that fully characterize the subjects in $\mathbf{v}$. For notational simplicity we henceforth write F to denote $F(\mathbf{v})$, and we let the dependence on $\mathbf{v}$ be implied.

III. The role of interference on the reliability of SBSs: Analyzing the probability of error for an authentication group

Towards evaluating the overall probability of identification error, we first establish the probability of error for a given set (authentication group) $\mathbf{v}$. We note the two characteristic extreme instances of $F(\mathbf{v}) = N$ and $F(\mathbf{v}) = 1$. In the first case, the random N-tuple $\mathbf{v}$ over which authentication will take place, happens to be such that each subject in $\mathbf{v}$ belongs to a different category, in which case none of the subjects interferes with another subject's identification. On the other hand, the second case corresponds to the (unfortunate) realizations of $\mathbf{v}$ where all subjects in $\mathbf{v}$ fall under the same category (all subjects in $\mathbf{v}$ happen to share the same features), and where authentication is highly unreliable.

Before proceeding with the analysis, we briefly define some notation. First we let $P_\phi,\ \phi \in \Phi$, denote the probability of incorrectly identifying a subject from S_ϕ, and we adopt for now the simplifying assumption that this probability be independent of the specific subject in S_ϕ. Without loss of generality, we also let $S_1, \cdots, S_F$ correspond to the $F(\mathbf{v}) = F$ non-empty categories, and note that $F \leq N$ since one subject can belong to just one category. Furthermore we let

$$S := \cup_{\phi=1}^{F} S_\phi$$

denote the set of subjects in $\mathbf{v}$ that can potentially be identified by the SBS endowed with Φ, and we note that $S = \cup_{\phi=1}^{\rho} S_\phi$. Also note that $|S_0| = N - |S|$, that $S_\phi \cap S_{\phi'} = \emptyset$ for $\phi' \neq \phi$, and that

$$|S| = \sum_{\phi=1}^{F} |S_\phi|.$$

We proceed to derive the error probability for any given $\mathbf{v}$.

Lemma 1: Let a subject be drawn uniformly at random from a randomly drawn N-tuple $\mathbf{v}$. Then the probability $P(\text{err}|\mathbf{v})$ of erroneously identifying that subject, is given by

$$P(\text{err}|\mathbf{v}) = 1 - \frac{F - \sum_{\phi=1}^{F} P_\phi}{N}, \tag{1}$$

where $F(\mathbf{v}) = F$ is the number of effective categories spanned by $\mathbf{v}$.

The following corollary holds for the interference limited case where errors due to feature estimation are ignored, i.e., where $P_\phi = 0$.

Corollary 2: For the same setting and measure as in Lemma 1, under the interference limited assumption, the probability of error $P(\text{err}|\mathbf{v})$ is given by

$$P(\text{err}|\mathbf{v}) = 1 - \frac{F}{N}, \tag{2}$$

for any $\mathbf{v}$ such that $F(\mathbf{v}) = F$.

The above reveals the somewhat surprising fact that, given N, the reliability of an SBS for identification of subjects in $\mathbf{v}$, is independent of the subjects' distribution $\mathbf{v}$ in the different categories, and instead only depends on F. As a result this reliability remains identical when employed over different N-tuples that fix F.

Proof of Lemma 1: See Appendix.

We proceed with a clarifying example.

Example *2:* Consider an SBS equipped with three features ($\rho = 3$), limited to (correctly) identifying dark hair, gray hair, and blond hair, i.e., $\Phi = \{$'dark hair' $= \phi_1$, 'gray hair' $= \phi_2$, 'blond hair' $= \phi_3\}$. Consider drawing at random, from a population corresponding to the residents of Nice, three N-tuples, with $N = 12$, each with a different subject categorization, as shown in Table I. Despite their different category distribution, the first two sets $\mathbf{v}_1$ and $\mathbf{v}_2$ introduce the same number of effective categories $F = 3$, and hence the same probability of erroneous detection $P(\text{err}|\mathbf{v}_1) = P(\text{err}|\mathbf{v}_2) = 3/4$ (averaged over the subjects in each set). On the other hand for $\mathbf{v}_3$ with $F = 2$, the probability of error increases to $P(\text{err}|\mathbf{v}_3) = 5/6$.

TABLE I
ILLUSTRATION OF EXAMPLE 2

	ϕ_1	ϕ_2	ϕ_3	F	$P(\text{err}\|\mathbf{v})$
$\mathbf{v}_1$	10	1	1	3	3/4
$\mathbf{v}_2$	4	4	4	3	3/4
$\mathbf{v}_3$	10	2	0	2	5/6

Up to now the result corresponded to the case of specific realizations of $\mathbf{v}$, where we saw that the probability of error for each realization of length N, was a function only of the realization of $F(\mathbf{v})$ which was a random variable describing the number of categories spanned by the specific group $\mathbf{v}$. We now proceed to average over all such realizations $\mathbf{v}$, and describe the overall probability of error. This analysis is better suited to evaluate an ensemble of distributed SBSs deployed over a large population. *We henceforth focus on the interference limited setting*[1] *i.e., we make the simplifying assumption that* $P_\phi = 0,\ \phi > 0$.

Lemma 3: The probability of error averaged over all N-tuples $\mathbf{v}$ randomly drawn from a sufficiently large population, is given by

$$\mathbb{E}_{\mathbf{v}}[P(\text{err}|\mathbf{v})] = 1 - \frac{\mathbb{E}_{\mathbf{v}}[F(\mathbf{v})]}{N}, \tag{3}$$

and is dependent only on the first order statistics of F.

Proof: The proof follows directly from Lemma 1. □

[1]We here note that with increasing ρ, the probability of erroneous identification is, in real systems, expected to increase. This will be considered in future work.

An illustration of the probability of error for a real distribution (Feret [4]) is given in Figure 1. An example follows,

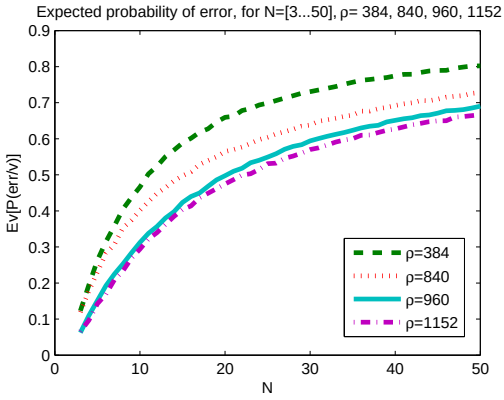

Fig. 1. $\mathbb{E}_{\mathbf{v}}[P(\text{err}|\mathbf{v})]$ for $\rho = 1152, 960, 840, 384$, $N \in [3, 4, ..., 50]$. Data from Feret [4].

related to the above.

Example *3:* Consider the case where the city of Nice installs throughout the city a number of independent SBSs[2] and is interested to know the average reliability that these systems will jointly provide, over a period of two months[3]. The result in Lemma 3 gives the general expression of the average reliability that is jointly provided by the distributed SBSs, indexed by N, for all N. Indexing by N simply means that the average is taken over all cases where authentication is related to a random set $\mathbf{v}$ of size N.

We now proceed to establish the statistical behavior of F, including the mean $\mathbb{E}[F]$.

IV. Analysis of Interference Patterns in SBSs: Establishing the Statistical Behavior of F

Given ρ and N, we are interested in establishing the probability $P(F)$ that a randomly drawn N-tuple of people will have F active categories out of a total of $\min(\rho, N)$ possible active categories[4]. We here accept the simplifying assumption of *uniform distribution* of the observed subjects over the categories ρ, i.e., that

$$P(\mathbf{v}(i) \in S_\phi) = \frac{1}{\rho},\ \forall \phi \in \Phi,\ i \leq N. \tag{4}$$

We also accept that $N < \rho$. The following then holds.

Lemma 4: Given ρ and N, and under the uniformity assumption, the distribution of F is described by

$$P(F) = \frac{F^{N-F}}{(\rho - F)!(N - F)! \sum_{i=1}^{N} \frac{i^{N-i}}{(N-i)!(\rho-i)!}}, \tag{5}$$

where F can take values between 1 and N.

Proof of Lemma 4: See Appendix.

[2]Independence follows from the assumption that the different SBSs are placed sufficiently far apart.

[3]In this example it is assumed that the number of independent SBSs and the time period are sufficiently large to jointly allow for ergodicity.

[4]Clarifying example: What is the statistical behavior of F that is encountered by a distributed set of SBSs in the city of Nice?

Example *4:* Consider the case where $\rho = 9, N = 5, F = 3$. Then the cardinality of the set of all possible N-tuples that span $F = 3$ effective categories, is given by the product of the following three terms.

- The first term is $(\rho \cdot (\rho-1) \cdots (\rho - F + 1)) = \frac{\rho!}{(\rho-F)!} = 9 \cdot 8 \cdot 7 = 504$ which describes the number of ways one can pick which $F = 3$ categories will be filled.
- Having picked these $F = 3$ categories, the second term is $(N \cdot (N-1) \cdots (N - F + 1)) = \frac{N!}{(N-F)!} = 5 \cdot 4 \cdot 3 = 60$, which describes the number of ways one can place exactly one subject in each of these picked categories.
- We are now left with $N - F = 2$ subjects, that can be associated freely to any of the $F = 3$ specific picked categories. Hence the third term is $F^{N-F} = 3^2 = 9$ corresponding to the cardinality of $\{1, 2, \cdots, F\}^{N-F}$.

Motivated by Lemma 3, we now proceed to describe the first order statistics of F. The proof is direct.

Lemma 5: Under the uniformity assumption, the mean of F is given by

$$\mathbb{E}_{\mathbf{v}}[F(\mathbf{v})] = \sum_{F=1}^{N} FP(F) = \sum_{F=1}^{N} \frac{\frac{F^{N-F+1}}{(\rho-F)!(N-F)!}}{\sum_{i=1}^{N} \frac{i^{N-i}}{(N-i)!(\rho-i)!}}. \quad (6)$$

Remark 1: The event of no interference corresponds to the case where $F = N$. Decreasing values of $\frac{F}{N}$ imply higher degrees of interference. An increasing ρ also results in reduced interference.

Related cases are plotted in Figure 2. A graphical representation of $E(F)$ for real-life data (Feret [4]) can be seen in Figure 3.

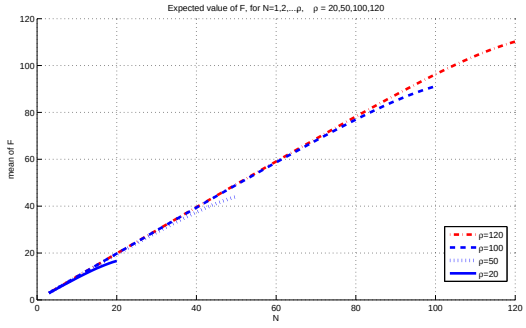

Fig. 2. $\mathbb{E}_{\mathbf{v}}[F]$ for $\rho = 20, 50, 100, 120$, $N \in [3, 4, ..., \rho]$. We note that for ρ sufficiently larger than N, then $\mathbb{E}_{\mathbf{v}}[F] \approx N$.

Finally, directly from the above, we have the following.

Theorem 6: In the described operational setting of interest, under the interference limited and uniformity assumptions, the probability of error averaged over all possible N-tuples $\mathbf{v}$, that is provided by an SBS endowed with ρ categories, is given by

$$P_{av}(\text{err}) = 1 - \frac{F^{N-F+1}}{(\rho-F)!(N-F)!N\sum_{i=1}^{N} \frac{i^{N-i}}{(N-i)!(\rho-i)!}}. \quad (7)$$

Proof of Theorem 6: The proof is direct from Lemma 3 and from (6). □

Related examples are plotted in Figure 4.

Furthermore we have the following.

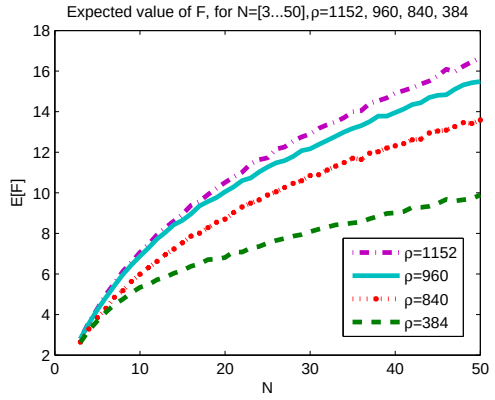

Fig. 3. $\mathbb{E}(\mathbb{F})$ for $\rho = 1152, 960, 840$ and 384, $N \in [3, 4, ..., 50]$ for the real data of Feret [4].

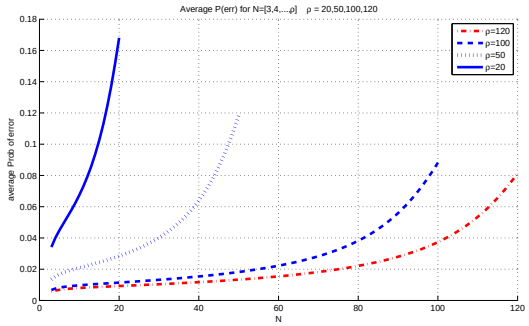

Fig. 4. $\mathbb{E}_{\mathbf{v}}[P(\text{err})]$ for $\rho = 20, 50, 100, 120$, $N \in [3, 4, ..., \rho]$.

Corollary 7: Under the uniformity assumption, the probability that interference exists, is given by

$$1 - P(N) = 1 - (\rho - N)! \sum_{i=1}^{N} \frac{i^{N-i}}{(N-i)!(\rho-i)!}. \quad (8)$$

Related examples are plotted in Figure 5.

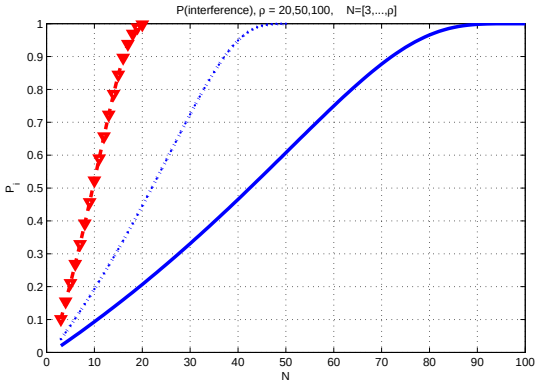

Fig. 5. $P(\text{interference exists})$ for $\rho = 20, 50, 100$, $N \in [3, 4, ..., \rho]$.

Example *5:* Given a group of $N = 10$ subjects, given an SBS with $\rho = 365$ categories, and under the uniformity assumption, the probability that interference exists is $1 - P(N) = 0.117$. Furthermore the same probability exceeds 0.5 for $N \geq 23$ subjects.

V. Asymptotic bounds on subject interference

In this section we seek to gain insight on the role of increasing resources (increasing ρ) in reducing the subject interference experienced by an SBS. Specifically we seek to gain insight on the following practical question: if more funds are spent towards increasing the quality of an SBS by increasing ρ, then what reliability gains do we expect to see? This question is only partially answered here, but some insight is provided in the form of bounds on the different subject-interference patterns seen by an SBS. The asymptotic bounds simplify the hard to manipulate results of Lemma 4 and Theorem 6, and provide insightful interpretations. A motivating example is presented before the result.

Example *6:* Consider an SBS operating in the city of Berlin, where for a specific N, this system allows for a certain average reliability. Now the city of Berlin is ready to allocate further funds, which can be applied towards doubling the number of categories ρ that the system can identify. Such an increase can come about, for example, by increasing the number and quality of sensors, which can now better identify more soft-biometric traits. The natural question to ask is how this extra funding will help improve the system? The bounds, when tight, suggest that doubling ρ, will result in a doubly exponential reduction in the probability that a specific degree of interference will occur.

Further clarifying examples that motivate this approach are given in Section V-A.

The following describes the result.

Lemma 8: Let

$$r := \lim_{\rho\to\infty} \frac{N}{\rho}, \tag{9}$$

define the *relative throughput* of a soft biometrics system, and let $F := fN,\ 0 \le f \le 1$. Then the asymptotic behavior of $P(F)$ is bounded as

$$-\lim_{\rho\to\infty} \frac{1}{\rho \log \rho} \log P(f) \ge 2 - r(1+f). \tag{10}$$

Proof of Lemma 8: See Appendix.

A. Interpretation of bounds

Lemma 8 bounds the statistical behavior of $P(F)$ in the high ρ regime. To gain intuition we compare two cases corresponding to two different relative-throughput regimes. In the first case we ask that N is close to ρ, corresponding to the highest relative-throughput of $r = 1$, and directly get from (10) that $d(r, f) := 2 - r(1 + f) = d(1, f) = 1 - f,\ 0 < f < 1$. In the second case we reduce the relative-throughput to correspond to the case where N is approximately half of ρ ($r = 1/2$), which in turn gives $d(r, f) = d(\frac{1}{2}, f) = \frac{3}{2} - \frac{f}{2},\ 0 < f < 1/2$. As expected $d(\frac{1}{2}, f) > d(1, f),\ \forall f \le \frac{1}{2}$.

Towards gaining further insight, let us use this same example to shed some light on how Lemma 8 succinctly quantifies the increase in the probability that a certain amount of interference will occur, for a given increase in the relative-throughput of the soft biometrics system. To see this, consider the case where there is a deviation away from the typical $f = r$ by some small *fixed* ϵ, to a new $f = r - \epsilon$, and note that the value of ϵ defines the extend of the interference[5], because a larger ϵ implies a smaller f, and thus a reduced F for the same N. In the high relative-throughput case of our example, we have that $f = r - \epsilon = 1 - \epsilon$, and thus that $d(1, 1 - \epsilon) = \epsilon$, which implies that the probability of such deviation (and of the corresponding interference) is in the order of $\rho^{-\rho d(1,1-\epsilon)} = \rho^{-\rho\epsilon}$. On the other hand, in the lower relative-throughput case where $f = r - \epsilon = \frac{1}{2} - \epsilon$, we have that $d(\frac{1}{2}, \frac{1}{2} - \epsilon) = \frac{5}{4} + \frac{\epsilon}{2}$, which implies that the probability of the same deviation in the lower throughput setting is in the order of $\rho^{-\rho d(\frac{1}{2},\frac{1}{2}-\epsilon)} = \rho^{-\rho(\frac{5}{4}+\frac{\epsilon}{2})} << \rho^{-\rho\epsilon}$. In other words the bound in Lemma 8 implies that, a reduction of the relative-throughput from its maximal value of $N/\rho \approx 1$ to a sufficiently smaller $N/\rho \approx \frac{1}{2}$, for high enough ρ, results in a substantial and exponential reduction in the probability of interference, from $P(r = 1) \approx \rho^{-\rho\epsilon}$ to $P(r = \frac{1}{2}) \approx \rho^{-\rho(\frac{5}{4}+\frac{\epsilon}{2})}$.

VI. Conclusions

The work explored the use of multi-trait SBSs for human identification, studying analytically the relationship between an authentication group $\mathbf{v}$, its size N, the featured categories ρ, and the effective categories F.

In the first part of the paper we showed that in the interference limited setting, for a given randomly chosen authentication group $\mathbf{v}$, of a given size N, the reliability of authentication (averaged over the subjects in $\mathbf{v}$) is a function only of the number of non-empty categories $F(\mathbf{v})$.

In the second part we provided statistical analysis of this reliability, over large populations. The latter part provided bounds that, in the interference limited setting suggest an *exponential* reduction in the probability of interference patterns, as a result of a *linear* increase in ρ.

VII. Appendix: Proofs

A. Proof of Lemma 1

Let $\widehat{\phi}$ denote the estimated category and let $P(S_\phi)$ denote the probability that the chosen subject belongs to category indexed by ϕ, $\phi = 0, 1, \cdots, F$. Then we have

$$\begin{aligned}
P(\text{err}|F) &= \sum_{\phi=0}^{F} P(S_\phi, \widehat{\phi} = \phi) P(\text{err}|S_\phi, \widehat{\phi} = \phi) \\
&\quad + \sum_{\phi=0}^{F} P(S_\phi, \widehat{\phi} \neq \phi) P(\text{err}|S_\phi, \widehat{\phi} \neq \phi) \\
&\overset{(a)}{=} \sum_{\phi=0}^{F} P(S_\phi) P(\widehat{\phi} = \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} = \phi) \\
&\quad + \sum_{\phi=0}^{F} P(S_\phi) P(\widehat{\phi} \neq \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} \neq \phi) \\
&\overset{(b)}{=} \frac{N - |S|}{N} + \sum_{\phi=1}^{F} P(S_\phi) P(\widehat{\phi} = \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} = \phi) \\
&\quad + \sum_{\phi=1}^{F} P(S_\phi) P(\widehat{\phi} \neq \phi|S_\phi) P(\text{err}|S_\phi, \widehat{\phi} \neq \phi). \qquad (11)
\end{aligned}$$

[5]Note that interference may occur only if $\epsilon > 0$.

Hence

$$\begin{aligned} &P(\text{err}|F) \\ &\overset{(c)}{=} \frac{N-|S|}{N} + \sum_{\phi=1}^{F}\big(\frac{|S_\phi|}{N}(1-P_\phi)\frac{|S_\phi|-1}{|S_\phi|} + \frac{|S_\phi|}{N}P_\phi\big) \\ &= \frac{N-|S|}{N} + \sum_{\phi=1}^{F}\big(|S_\phi| - 1 - P_\phi|S_\phi| + P_\phi + P_\phi|S_\phi|\big) \end{aligned} \tag{12}$$

which gives

$$P(\text{err}|F) = 1 - \frac{|S|}{N} + \frac{1}{N}\sum_{\phi=1}^{F}\big(|S_\phi| - 1 + P_\phi\big) \tag{13}$$

$$\overset{(d)}{=} 1 - \frac{F - \sum_{\phi=1}^{F} P_\phi}{N}. \tag{14}$$

In the above (a) is due to Bayes rule, (b) considers that

$$P(\text{err}|S_0, \widehat{\phi} = 0) = P(\text{err}|S_0, \widehat{\phi} \neq 0) = 1$$

and that

$$\begin{aligned} &P(S_0, \widehat{\phi} = 0)P(\text{err}|S_0, \widehat{\phi} = 0) \\ &\qquad + P(S_0, \widehat{\phi} \neq 0)P(\text{err}|S_0, \widehat{\phi} \neq 0) \\ &= P(S_0, \widehat{\phi} = 0)\cdot 1 + P(S_0, \widehat{\phi} \neq 0)\cdot 1 = P(S_0) = \frac{N-|S|}{N}, \end{aligned}$$

(c) considers that $P(S_\phi) = \frac{|S|}{N}$, that $P(\widehat{\phi} = \phi|S_\phi) = 1 - P_\phi$, that $P(\text{err}|S_\phi, \widehat{\phi} \neq \phi) = 1$, and that

$$P(\text{err}|S_\phi, \widehat{\phi} = \phi) = \frac{|S_\phi| - 1}{|S_\phi|},$$

and finally (d) considers that $\sum_{\phi=1}^{F}|S_\phi| = |S|$.

□

B. Proof of Lemma 4

Let C_F be the total number of N-tuples $\mathbf{v}$ that introduce F effective feature categories. Then

$$C_F = \frac{\rho!}{(\rho - F)!}\frac{N!}{(N-F)!}F^{N-F} \tag{15}$$

where the first term $\frac{\rho!}{(\rho-F)!}$ describes the total number of ways F categories can be chosen to host subjects, the second term $\frac{N!}{(N-F)!}$ describes the total number of ways F initial people, out of N people, can be chosen to fill these F categories, and where the third term F^{N-F} describes the total number of ways the F effective categories can be freely associated to the rest $N - F$ subjects. Finally we note that

$$P(F) = \frac{C_F}{\sum_{i=1}^{N} C_i},$$

which completes the proof. □

C. Proof of Lemma 8

Recall from (5) that

$$P(F) = \frac{F^{N-F}}{(\rho-F)!(N-F)!\sum_{i=1}^{N} i^{N-i}\big((N-i)!(\rho-i)!\big)^{-1}}, \tag{16}$$

and note that

$$\sum_{i=1}^{N} i^{N-i}\big((N-i)!(\rho-i)!\big)^{-1} \geq (\rho - N)!$$

corresponding to the Nth summand $(i = N)$, and corresponding to the fact that all summands are non-negative. As a result

$$P(F) \leq \frac{F^{N-F}}{(\rho-F)!(N-F)!(\rho-N)!}.$$

Using Stirling's approximation [11] that holds in the asymptotically high ρ setting of interest, we have

$$P(F) \dot{\leq} \frac{F^{N-F}}{(\rho-F)^{\rho-F}(N-F)^{N-F}(\rho-N)^{\rho-N}e^{-(2\rho-2F)}}, \tag{17}$$

and as a result

$$\begin{aligned} &P(f) \\ &\dot{\leq} \frac{(fr\rho)^{r\rho(1-f)}}{(\rho - fr\rho)^{\rho-fr\rho}(r\rho - fr\rho)^{r\rho-fr\rho}(\rho - r\rho)^{\rho-r\rho}e^{2\rho(1+fr)}} \\ &= \frac{\rho^{r\rho(1-f)}\rho^{-\rho(1-fr)}}{(fr)^{-r\rho(1-f)}(1-fr)^{\rho(1-fr)}} \\ &\quad \cdot \frac{\rho^{-\rho r(1-f)}\rho^{-\rho(1-r)}}{(r-fr)^{\rho r(1-f)}(1-r)^{\rho(1-r)}e^{2\rho(1+fr)}}. \end{aligned} \tag{18}$$

In the above we use $\doteq$ to denote *exponential equality*, where

$$f \doteq \rho^{-\rho B} \iff -\lim_{\rho\to\infty}\frac{\log f}{\rho\log\rho} = B, \tag{19}$$

with $\dot{\leq}, \dot{\geq}$ being similarly defined. The result immediately follows. □

REFERENCES

[1] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," *in Proc ICBA,* 2004, pp. 731-738.

[2] A. K. Jain, S. C. Dass, and K. Nandakumar, "Can soft biometric traits assist user recognition?," *in Proc. of SPIE,* 2004, vol. 5404, pp. 561-572.

[3] Henry T.F. Rhodes (1956), *Alphonse Bertillon: Father of Scientific Detection,* Abelard-Schuman, New York, Greenwood Press.

[4] http://face.nist.gov/colorferet/

[5] A. Dantcheva, J.-L. Dugelay, and P. Elia, "Person recognition using a bag of facial soft biometrics (BoFSB)," *in Proc. MMSP,* 2010.

[6] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft-biometrics: unconstrained authentication in a surveillance environment," *in Proc. DICTA,* 2009, pp. 196-203.

[7] G.L. Marcialis, F. Roli, and D. Muntoni, "Group-specific face verification using soft biometrics," *Journal of Visual Languages and Computing,* 2009, vol. 20, pp. 101-109.

[8] K. Moustakas, D. Tzovaras, and G. Stavropoulos "Gait recognition using geometric features and soft biometrics," *Signal Processing Letters,* 2010, vol. 17, pp. 367-370.

[9] S. Samangooei, M. Nixon, and B. Guo, "The use of semantic human description as a soft biometric," *in Proc. BTAS,* 2008.

[10] D. Meltem, G. Kshitiz, and G. Sadiye, "Automated person categorization for video surveillance using soft biometrics," *in Proc. SPIE,* 2010.

[11] M. Abramowitz, and I. Stegun, (2002), *Handbook of Mathematical Functions*, Dover Publications, New York, USA.