Group Membership Verification with Privacy: Sparse or Dense?

Marzieh Gheisari, Teddy Furon, Laurent Amsaleg

Group Membership Verification with Privacy: Sparse or Dense?

Marzieh Gheisari
Univ Rennes, Inria, CNRS, IRISA
France

Teddy Furon
Univ Rennes, Inria, CNRS, IRISA
France

Laurent Amsaleg
Univ Rennes, Inria, CNRS, IRISA
France

***Abstract*—Group membership verification checks if a biometric trait corresponds to one member of a group without revealing the identity of that member. Recent contributions provide privacy for group membership protocols through the joint use of two mechanisms: quantizing templates into discrete embeddings, and aggregating several templates into one group representation.**

However, this scheme has one drawback: the data structure representing the group has a limited size and cannot recognize noisy query when many templates are aggregated. Moreover, the sparsity of the embeddings seemingly plays a crucial role on the performance verification.

This paper proposes a mathematical model for group membership verification allowing to reveal the impact of sparsity on both security, compactness, and verification performances. This models bridges the gap towards a Bloom filter robust to noisy queries. It shows that a dense solution is more competitive unless the queries are almost noiseless.

I. Introduction

Group membership verification is a procedure checking whether an item or an individual is a member of a group. If membership is positively established, then an access to some ressources (a building, a file, ...) is granted; otherwise the access is refused. This paper focuses on *privacy preserving* group membership verification procedures where members must be distinguished from non-members, but where the members of a group should not be distinguished one another.

To this aim, a few recent contributions have proposed to rely on the aggregation and the embedding of several distinctive templates into a unique and compact high dimensional feature representing the members of a group [1], [2]. It has been demonstrated that this allows a good assessment of the membership property at test time. It has also been shown that this provides privacy and security. Privacy is enforced because it is impossible to infer from the aggregated feature which original distinctive template matches the one used to probe the system. Security is preserved since nothing meaningful leaks from embedded data [3], [4].

[1] and [2], however face severe limitations. Basically, it seems impossible to create features representing groups having many members. In this case, the probability to identify true positives vanishes and the false negative rate grows accordingly. Furthermore, the robustness of the matching procedure fades and becomes unable to absorb even the smallest amount of noise that inherently differentiate the enrolled template of one member and the template captured at query time for this same member. In contrast, features representing only few group members are robust to noise and cause almost no false negatives. A detailled analysis of [1] and [2] suggests that these limitations originate from the sparsity level of the features representing group members.

This paper investigates the impact of the sparsity level of the high dimensional features representing group members on the quality of (true positive) matches and on their robustness to noise. It shows it is possible to trade compactness and sparsity for better security or better verification performance.

Sect. II first considers the aggregation of discrete random sequences, and models this compromise with information theoretical tools. Sect. III applies this viewpoint to binary random sequences and shows that the noise on the query has an impact depending on the sparsity of the sequences. Sect. IV bridges the gap between the templates, *i.e.* real d-dimensional vectors, and the discrete sequences considered in the previous sections. Sect. V gathers the experimental results for a group membership verification based on faces.

II. Discrete Sequences

This section considers the problem of creating a representation $\mathbf{Y}$ of a group of n sequences $\{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$, whose use is to test whether a query sequence $\mathbf{Q}$ is a noisy version of one of these n original sequences. This test is done at query time when the original sequences are no longer available and all that remains is the representation $\mathbf{Y}$.

The sequences are elements of $\mathcal{X}^m$ where $\mathcal{X}$ is a finite alphabet of cardinality $|\mathcal{X}|$, say $\mathcal{X} := \{0, 1, \ldots, |\mathcal{X}| - 1\}$. The sequence follows a statistical model giving a central role to the symbol 0. The symbols of the sequences are independent and identically distributed with

$$\mathbb{P}(X = s) = \begin{cases} 1 - p(|\mathcal{X}| - 1) & \text{if } s = 0 \\ p & \text{otherwise} \end{cases} \quad (1)$$

for $p \in (0, 1/|\mathcal{X}|]$. Sparsity means that probability p is small, density means that p is close to $1/|\mathcal{X}|$ so that X is uniformly distributed over $\mathcal{X}$.

WIFS'2019, December, 9-12, 2019, Delft, Netherlands. 978-1-7281-3217-4/19/$31.00 ©2019 European Union.

A. Structure of the group representation

We impose the following conditions on the aggregation $\mathsf{a}(\cdot)$ computing the group representation $\mathbf{Y} = \mathsf{a}(\mathbf{X}_1, \dots, \mathbf{X}_n)$:

- $\mathbf{Y}$ is a discrete sequence of the same length $\mathbf{Y} \in \mathcal{Y}^m$,
- Symbol $Y(i)$ only depends on symbols $\{X_1(i), \dots, X_n(i)\}$,
- The same aggregation is made index-wise: with abuse of notation, $Y(i) = \mathsf{a}(X_1(i), \dots, X_n(i))$, $\forall i \in [m]$,
- $Y(i)$ does not depend on any ordering of the set $\{X_1(i), \dots, X_n(i)\}$,

These requirements are well known in traitor tracing and group testing as they usually model the collusion attack or the test results over groups. Here, they simplify the analysis reducing the problem to a single letter formulation where index i is dropped involving symbols $\{X_1, \dots, X_n\}$, Y and Q.

These conditions motivate a 2-stage construction. The first stage computes the type (a.k.a. histogram or tally) T of the symbols $\{X_1, \dots, X_n\}$. Denote by $\mathcal{T}_{|\mathcal{X}|,n}$ the set of possible type values. Its cardinality equals $|\mathcal{T}_{|\mathcal{X}|,n}| = \binom{n+|\mathcal{X}|-1}{|\mathcal{X}|-1}$ which might be too big. The second stage applies a surjective function $\mathsf{r} : \mathcal{T}_{|\mathcal{X}|,n} \to \mathcal{Y}$, where $\mathcal{Y}$ is a much smaller set.

B. Noisy query

At enrollment time, the system receives n sequences, aggregates them into the compact representation $\mathbf{Y}$, and then forgets the n sequences. At query time, the system receives a new sequence $\mathbf{Q}$ conforming with one of the following hypotheses:

- $\mathcal{H}_1$: $\mathbf{Q}$ is a noisy version of one of the enrolled sequences. Without loss of generality, $\mathbf{Q} = \mathbf{X}_1 + \mathbf{N}$.
- $\mathcal{H}_0$: $\mathbf{Q} = \mathbf{X}_0 + \mathbf{N}$, where $\mathbf{X}_0$ shares the same statistical model but it is independent of $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$.

We model the source of noise (due to different acquisition conditions) by a discrete communication channel. It is defined by function $\mathsf{W} : \mathcal{X} \times \mathcal{X} \to [0,1]$ with $\mathsf{W}(q|x) := \mathbb{P}(Q = q|X = x)$. We impose some symmetry w.r.t. the symbol 0: $\mathsf{W}(s|0) = \eta_0$ and $\mathsf{W}(0|s) = \eta_1$, $\forall s \in \mathcal{X}\backslash\{0\}$.

At query time, the system computes a score $S = \mathsf{s}(\mathbf{Q}, \mathbf{Y})$ and compares to a threshold: hypothesis $\mathcal{H}_1$ is deemed true if $S \geq \tau$. This test leads to two probabilities of error:

- $P_{\mathsf{fp}}(n,m)$ is the probability of false positive: $P_{\mathsf{fp}}(n,m) := \mathbb{P}(S \geq \tau|\mathcal{H}_0)$.
- $P_{\mathsf{fn}}(n,m)$ is the probability of false negative: $P_{\mathsf{fn}}(n,m) := \mathbb{P}(S < \tau|\mathcal{H}_1)$.

The emphasis on (n,m) is natural. It is expected that: i) the more sequences are aggregated, the less reliable the test is, ii) the longer the sequences are, the more reliable the test is.

C. Figures of merit $(\mathsf{C}, \mathsf{S}, \mathsf{V})$

The section presents three information theoretic quantities (expressed in nats) measuring the performances of the scheme. The first two depends on the statistical model of X (especially p) and the aggregation mechanism a. The last one depends moreover on the channel.

1) Compactness C*:* The compactness of the group representation is measured by the entropy $\mathsf{C} := H(Y)$. It roughly means that the number of typical sequences $\mathbf{Y}$ scales exponentially as $e^{mH(Y)}$, which can be theoretically compressed to the rate of $H(Y)$ nats per symbol.

2) Security S*:* We consider an insider aiming at disclosing one of the n enrolled sequences. Observing the group representation $\mathbf{Y}$, its uncertainty is measured by the equivocation $\mathsf{S} := H(X|Y)$. This means that the insider does not know which of the $e^{mH(X|Y)}$ typical sequences the enrolled sequences are.

3) Verification V*:* In our application, the requirement of utmost importance is to have a very small probability of false positive. We are interested in an asymptotical setup where $m \to +\infty$. This motivates the use of the false positive error exponent as a figure of merit:

$$E_{\mathsf{fp}}(n) := \lim_{m \to +\infty} -\frac{1}{m} \log P_{\mathsf{fp}}(n,m). \tag{2}$$

If $E_{\mathsf{fp}}(n) > 0$, it means that $P_{\mathsf{fp}}(n,m)$ exponentially vanishes as m becomes larger. The theory of test hypothesis shows that $E_{\mathsf{fp}}(n)$ is upper bounded by the mutual information $\mathsf{V} := I(Y;Q)$ where Q is a symbol of the query sequence, *i.e.* a noisy version of X_1. It means that the necessary length for achieving the requirement $P_{\mathsf{fp}}(n,m) < \epsilon$ is [5]

$$m \geq \frac{-\log \epsilon}{\mathsf{V}}. \tag{3}$$

D. Noiseless setup

The bigger V and S, the better the performance in terms of verifiability and security. Yet, they can not be both big at the same time. The noiseless case when the channel introduces no error and $Q = X$ simply illustrates the trade-off:

$$\mathsf{V} \leq \mathsf{C} \tag{4}$$

$$\mathsf{V} + \mathsf{S} = H(X), \tag{5}$$

with $H(X) = -\log p_0 + (1-p_0)\log\frac{p}{p_0}$ and $p_0 := \mathbb{P}(X = 0)$ (1). For a given $|\mathcal{X}|$, $H(X)$ is maximised by the dense solution: $H(X) \leq \log|\mathcal{X}|$ with equality for $p = 1/|\mathcal{X}|$.

III. Binary alphabet

This section explores the binary case where $\mathcal{X} = \{0,1\}$. We first set the surjection as the identity function s.t. $Y = T$. Then, the impact of the surjection is investigated.

A. Working with types

In the binary case, there are $n+1$ type values. There can be uniquely labelled by the number of symbols '1' in $\{X_1, \dots, X_n\}$, *i.e.* $T = \sum_{i=1}^n X_i \sim \mathcal{B}(n,p)$.

1) Verification: In the noiseless case, after some rewriting:

$$\mathsf{V} = \mathsf{h}(p) - \sum_{t=0}^{n} \mathbb{P}(T = t)\mathsf{h}\left(\frac{t}{n}\right), \tag{6}$$

with $\mathsf{h}(p) := -p\log(p) - (1-p)\log(1-p)$, the entropy of a Bernoulli r.v. $\mathcal{B}(p)$. If $p = 1/2$ and n is large:

$$\mathsf{V} = \frac{1}{2n} + o\left(\frac{1}{n}\right). \tag{7}$$

This is not the maximum of this quantity. For large n, the best option is to set

$$p = \frac{\alpha}{n}, \quad \mathsf{V} = \frac{\beta}{n} + o\left(\frac{1}{n}\right), \tag{8}$$

with $\alpha = 1.338$ and $\beta = 0.580$. This was proven in the totally different application of traitor tracing [6, Prop. 3.8].

This section outlines two setups: the dense setup where $p = 1/2$, and the sparse setup where p goes to 0 when more sequences are packed in the group representation. Both setups share the asymptotical property that $\mathsf{V} \approx \kappa/n$ for large n. According to (3), we can pack a big number n of sequences into one group representation provided that their length m scales proportionally to n.

2) Compactness: The figure of merit for compactness for types is just $\mathsf{C} = H(T)$ where T follows a binomial distribution: $T \sim \mathcal{B}(n, p)$. In the dense setup $p = 1/2$, the binomial distribution is approximated by a Gaussian distribution $\mathcal{N}(n/2; n/4)$ providing:

$$\mathsf{C} = \frac{1}{2}\log\left(\frac{\pi e n}{2}\right) + O\left(\frac{1}{n}\right). \tag{9}$$

In the sparse setup $p = \alpha/n$, the binomial distribution is approximated by a Poisson distribution $\mathcal{P}(\alpha)$ [7]:

$$\mathsf{C} \approx \alpha(1 - \log(\alpha)) + e^{-\alpha}\sum_{j=0}^{+\infty}\frac{\alpha^j \log(j!)}{j!}. \tag{10}$$

This shows that the types are not compact in the dense setup; It approximatively remains constant in the sparse setup.

3) Security: Thanks to (5), we only need to calculate $H(X) = \mathsf{h}(p)$. In the dense setup, $H(X) = \log(2)$ and S converges to $H(X)$ as n increases. Merging into a single representation protects an individual sequence. If sparse,

$$H(X) = \frac{\alpha}{n}\left(1 - \log\frac{\alpha}{n}\right) + o\left(\frac{1}{n}\right). \tag{11}$$

Therefore, S converges to zero as n increases, contrary to the dense setup. It might be more insightful to see that the ratio of uncertainties before and after observing T, *i.e.* $H(X)/H(X|T)$, converges to 1 in both cases. Merging does provide some security but sparsity is more detrimental.

B. Adding a surjection

The motivation of the surjection onto a smaller set $\mathcal{Y}$ is to bound C as $\mathsf{C} \leq \log|\mathcal{Y}|$, $\forall n$. The Markov chain $Q \to X_1 \to T \to Y$ imposes that $\mathsf{V} \leq I(T; Q)$. The surjection thus provoques a loss in verification as depicted in Fig. 1.

App. A shows that for $|\mathcal{Y}| = 2$, this loss is minimized for:

$$\mathsf{r}(t) = \begin{cases} 0 & \text{if } t < t_p \\ 1 & \text{otherwise} \end{cases} \tag{12}$$

where t_p is a threshold depending on p. In the dense setup, $t_p = n/2$ and the surjection corresponds to a majority vote collusion in traitor tracing (a threshold model in group testing). Hence, by [6, Prop. 3.4]:

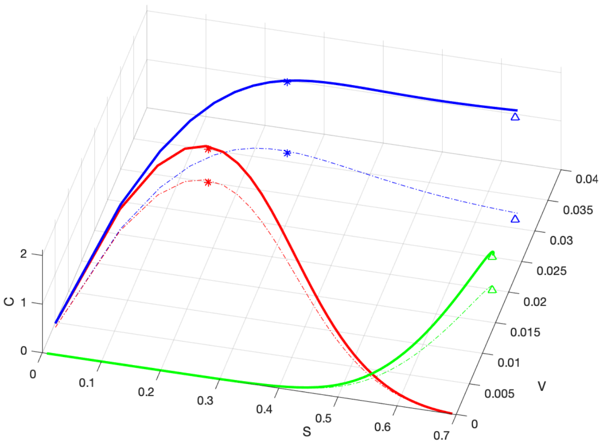

Fig. 1. The trade-off $(\mathsf{S}, \mathsf{V}, \mathsf{C})$ for $\mathcal{X} = \{0, 1\}$, $n = 16$, $Y = T$ (blue), $Y = \mathsf{r}(T)$ for 'All-1' (red) and majority vote (green). Dashed plot represents the projection onto $\mathsf{C} = 0$. Triangles and stars summarize results (7) to (14).

$$\mathsf{V} = \frac{1}{n\pi} + o\left(\frac{1}{n}\right). \tag{13}$$

In the sparse setup $t_p = 1$ which corresponds to an 'All-1' attack in traitor tracing (a the perfect model in group testing). Then the best option is to set $p = \log(2)/n$ and [6, Prop. 3.3]:

$$\mathsf{V} = \frac{(\log(2))^2}{n} + o\left(\frac{1}{n}\right). \tag{14}$$

From (3), the necessary length is $m \geq -n\log(\epsilon)/(\log(2))^2$.

The main property $\mathsf{V} \approx \kappa/n$ still holds but the surjection lowers κ from 0.5 to 0.32 (dense), from 0.84 to 0.48 (sparse). The sparse setup is still the best option w.r.t. V.

C. Relationship with the Bloom filter

A Bloom filter is a well-known data structure $\mathbf{Y} \in \{0, 1\}^m$ designed for set membership, embedding items to be enrolled into $\mathbf{Y}$ thanks to k hash functions. Its probability of false negative is exactly 0, whereas the probability of false positive is not null. The number of hash functions minimizing $P_{\mathsf{fp}}(n, m)$ is $k = \lfloor \log(2)m/n \rfloor$. Then, the necessary length to meet a required false positive level ϵ is $m \geq -n\log(\epsilon)/(\log(2))^2$.

These numbers show the connection with our scheme (14). At the enrollment phase, the hash functions indeed associate to the j-th item a binary sequence $\mathbf{X}_j$ indicating which bits of $\mathbf{Y}$ have to be set. This sequence is indeed sparse with $k/m \approx \log(2)/n$. The necessary length is the same. Indeed, the enrollment phase of a Bloom filter is nothing more than the 'All-1' surjection.

The only difference resides in the statistical model. There is at most k symbols '1' in sequence $\mathbf{X}_j$ whereas, in our model, that follows a binomial distribution $\mathcal{B}(m, p)$. Yet, asymptotically as $m \to \infty$, by some concentration phenomenon, the two models get similar. This explains why we end up with similar optimal parameters. Yet, the Bloom filter only works when the query object is exactly one enrolled item, whereas the next section shows that our scheme is robust to noise.

IV. Real vectors

This section deals with real vectors: n vectors to be enrolled $\{\vec{x}_1, \dots, \vec{x}_n\} \subset \mathbb{R}^d$, and the query vector $\vec{q} \in \mathbb{R}^d$. All have unit norm. An embedding mechanism $\mathsf{E} : \mathbb{R}^d \to \mathcal{X}^m$ makes the connection with the previous section. As in [8], this study models the embedding as a probabilistic function.

A. Binary embedding

For instance, for $\mathcal{X} = \{0, 1\}$, a popular embedding is:

$$X(i) = [\vec{x}^\top \vec{U}_i > \lambda_x], \forall i \in [m] \tag{15}$$

where $\vec{U}_i \overset{i.i.d.}{\sim} \mathcal{N}(\vec{0}_d, I_d)$. This in turn gives i.i.d. Bernoulli symbols $\{X(i)\}$ with $p = 1 - \Phi(\lambda_x)$ if $\|\vec{x}\| = 1$.

At the query time, the embedding mechanism uses the same random vectors but a different threshold:

$$Q(i) = [\vec{q}^\top \vec{U}_i > \lambda_q], \forall i \in [m]. \tag{16}$$

Under $\mathcal{H}_1$, suppose that $\vec{q}^\top \vec{x}_1 = c < 1$. This correlation defines the channel $X \to Q$ with the error rates:

$$\eta_0 = \mathbb{P}(\vec{q}^\top \vec{U} > \lambda_q | \vec{x}^\top \vec{U} \leq \lambda_x), \tag{17}$$

$$\eta_1 = \mathbb{P}(\vec{q}^\top \vec{U} \leq \lambda_q | \vec{x}^\top \vec{U} > \lambda_x). \tag{18}$$

The error rate η_0 has the expression (and similarly for η_1):

$$\eta_0 = 1 - \frac{1}{(1-p)\sqrt{2\pi}} \int_{-\infty}^{\lambda_x} \Phi\left(\frac{\lambda_q - cx}{\sqrt{1-c^2}}\right) e^{-\frac{x^2}{2}} dx. \tag{19}$$

B. Induced channel

For this embedding, the parameters $(\lambda_x, \lambda_q, c, d)$ for the vectors define the setup (p, η_0, η_1) for the sequences. It is a priori difficult to find the best tuning (λ_x, λ_q). For a fixed λ_x, η_0 decreases with λ_q while η_1 increases. App. B reveals that V is sensitive to η_0 especially with the 'All-1' surjection of the sparse solution. Fig. 2 shows indeed that the dense solution $(\lambda_x, \lambda_q) = (0, 0)$ is more robust, unless c is very close to 1. Here, we enforce a surjection (identity, All-1, or majority vote) and make a grid search to find the optimum (λ_x, λ_q) for a given c. It happens that these parameters are better set to 0, *i.e.* dense solution, for the identity and majority vote. As for the 'All-1' surjection, we observe that λ_x is s.t. $p \approx 1/n$ and λ_q is slightly bigger than λ_x to lower η_0. Yet, this sparse solution is not as good as the dense solution unless c is close to 1, *i.e.* the query vector is very close to the enrolled vector.

This observation holds only for the embedding function (15). Hashing functions less prone to error η_0 may exist.

V. Experimental work

We evaluate our scheme with face recognition. Face images are coming from LFW [9], CFP [10] and FEI [11] databases. For each dataset, N individuals are enrolled into random groups. There is the same number N_q of positive and negative (impostors) queries.

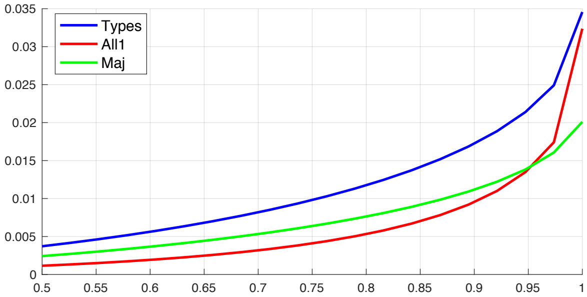

Fig. 2. V as a function of correlation c, $d = 256$, $n = 15$.

Labeled Faces in the Wild: These are pictures of celebrities in all sort of viewpoint and under an uncontrolled environment. We use pre-aligned LFW images. The enrollment set consists of $N = 1680$ individuals with at least two images in the LFW database. One random template of each individual is enrolled in the system, playing the role of $\vec{x}_i$. Some other $N_q = 263$ individuals were randomly picked in the database to play the role of impostors.

Celebrities in Frontal-Profile: These are frontal and profile views of celebrities taken in an uncontrolled environnement. We only use $N = 400$ frontal images enrolled in the system. The impostor set is a random selection of $N_q = 100$ other individuals.

Faculdade de Engenharia Industrial: The FEI database contains images in frontal view in a controlled environnement. We use pre-aligned images. There are 200 subjects with two frontal images (one with a neutral expression and the other with a smiling facial expression). The database is created by randomly sampling $N = 150$ individuals to be enrolled, and $N_q = 50$ impostors.

A. Experimental Setup

Face descriptors are obtained from a *pre-trained* network based on VGG-Face architecture followed by PCA [12] . FEI corresponds to the scenario of employees entering in a building with face recognition, whereas CFP is more difficult, and LFW even more difficult. To equalize the difficulty, we apply a dimension reduction (Probabilistic Principal Component Analysis [13]) to $d = 128$ (FEI), 256 (CFP), and 512 (LFW). The parameters of PPCA are learned on a different set of images, not on the enrolled templates and queries. The vectors are also L_2 normalized. With such post-processing, the average correlation between positive pairs equals 0.83 (FEI), 0.78 (CFP), and 0.68 (LFW) with a standard deviation of 0.01. Despite the dimension reduction, the hardest dataset is LFW and the easiest FEI.

In one simulation run, the enrollment phase makes random groups with the same number n of members. A user claims she/he belongs to group g. This claim is true under hypothesis $\mathcal{H}_1$ and false under hypothesis $\mathcal{H}_0$ (*i.e.* the user is an impostor). Her/his template is quantized to the sequence $\mathbf{Q}$, and $(\mathbf{Q}, g)$ is sent to the system, which compares $\mathbf{Q}$ to the group repre-

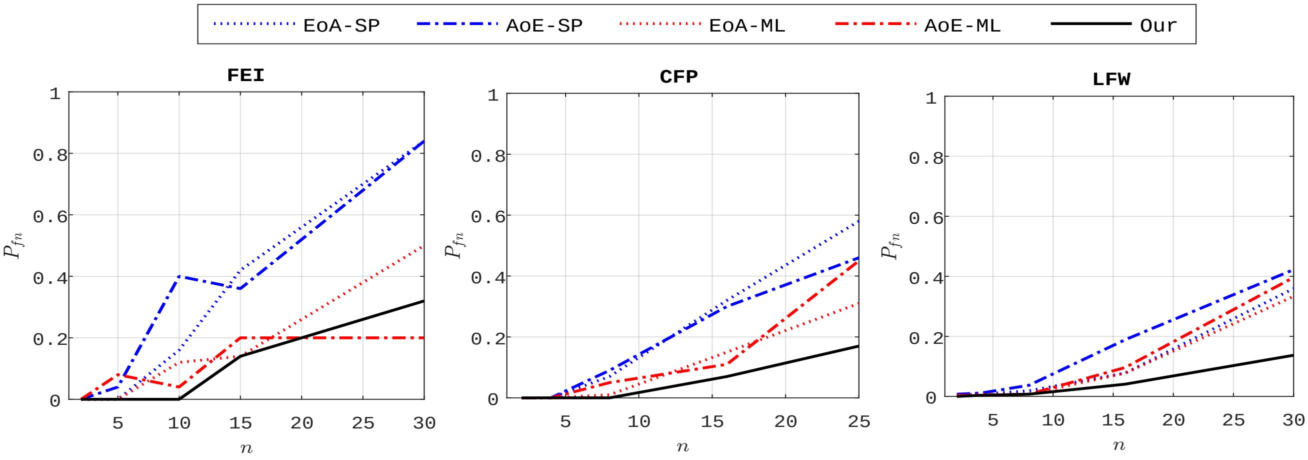

Fig. 3. Verification performance $P_{\mathsf{fn}}@P_{\mathsf{fp}} = 0.05$ *vs.* group size n for the baselines (see Sect. V-B) and our scheme.

sentation $\mathbf{Y}_g$. This is done for all impostors and all queries of enrolled people. One Monte-Carlo simulation is composed of 20 runs. The figure of merit is P_{fn} when $P_{\mathsf{fp}} = 0.05$.

B. Exp. #1: Comparison to the baselines

Our scheme is compared to the following baselines:

- EoA-SP and AoE-SP [1] (signal processing approach)
- EoA-ML and AoE-ML [2] (machine learning approach)

The drawback of these baselines is that the length m of the data structure is bounded. Here, it is set to maximum value, *i.e.* $m = d$ the dimension of templates.

Our scheme allows more freedom. Setting $m = 8 \times d$ produces a much bigger representation. It is not surprising that our scheme is better than the baselines. Fig. 3 validates our motivation to get rid off the drawback of the baselines with limited m, to achieve better verification performance. These results are obtained with the dense solution. Indeed, despite all our efforts, we could not achieve better results with the sparse solution. This confirms the lesson learnt from Fig. 2: the dense solution outperforms the sparse solution when the average correlation between positive pairs is lower than 0.95.

The improvement is also better as the size of groups increases. We explain this by the use of the types, *i.e.* $Y = T$. Equation (9) shows that C increases with n for the dense solution, compensating for aggregating more templates.

C. Exp. #2: Reducing the size of the group representation

There are two ways for reducing the size of the group representation. The first means is to decrease m, the second means is to lower C thanks to a surjection. Sect. III-B presented optimal surjections from $\mathcal{T}_{2,n}$ to $\mathcal{Y} = \{0, 1\}$. We found experimentally good surjections to sets $\mathcal{Y}$ for $|\mathcal{Y}| \in \{3, 4, 8\}$.

This is done according to the following heuristic. Starting from $\mathcal{T}_{2,n}$, we iteratively decrease the size of $\mathcal{Y}$ by one. This amounts to merge two symbols of $\mathcal{Y}$. By brute force, we analyse all the pairs of symbols measuring the loss in V induced by their merging. By merging the best pair, we decrease the number of symbols in $\mathcal{Y}$ by one. This process is iterated until the targeted size of $\mathcal{Y}$ is achieved. This heuristic is not optimal, but it is tractable. Fig. 4 compares these two means. Employing a coarser surjection is slightly better in terms of verification performances.

D. Unexpected results

We have argued that FEI $<$ CFP $<$ LFW in terms of difficulty due to the opposite ordering of the datasets typical correlation c between positive pairs. Eq. (19) shows that a lower c produces a higher η_0 (and η_1), whence a lower V. In Fig. 3, the experimental results contradict this intuition.

This may be explained by the Signal to Noise Ratio at the template level. We define it as c^2/v_0 where c is the average correlation for positive pairs and v_0 is the variance of this correlation for negative pairs. If a negative query is uniformly distributed over the hypersphere, then its correlation with an enrolled template is approximatively distributed as a centered Gaussian distribution with variance $v_0 = 1/d$.

Yet, d has no impact on p, η_0, and η_1. We suppose that its impact is tangible on the entropy of the template vectors. Sect. II assumes that the enrolled sequences are statistically independent. This assumption is not granted with the embedding of Sect. IV. Yet, a bigger d favors the independence (or at least the decorrelation) between real template vectors.

VI. Conclusion

Our theoretical study justifies that the dense setup is more interesting in terms of verification performance V and security level S unless we are operating in the high-SNR regime where the positive queries are very well correlated with the enrolled templates. This statement holds for any embedding, yet some are certainly more suited than others depending on d, c, and the geometrical relationship among positive pairs.

Acknowledgment

This work is supported by the project CHIST-ERA ID_IOT 20CH21 167534.

Appendix

Let us first explain how V is computed. Denote $P_i(q,y) := \mathbb{P}(Q=q, Y=y|\mathcal{H}_i)$ and channel $W(q|x) := \mathbb{P}(Q=q|X=x)$, $\forall y \in \mathcal{Y}, q \in \mathcal{X}$ and $i \in \{0,1\}$. Then,

$$\mathsf{V} = \sum_{q,y} P_1(q,y) \log \frac{P_1(q,y)}{P_0(q,y)}, \tag{20}$$

with $P_0(q,y) = \mathbb{P}(Q=q)\mathbb{P}(Y=y)$ and

$$P_1(q,y) = \sum_{x\in\mathcal{X}} \mathbb{P}(Y=y, X=x) W(q|x). \tag{21}$$

A. Surjection to $\mathcal{Y} = \{0,1\}$

We assume here the noiseless setup allowing to write $\mathbb{P}(Y=y, X=x)$ as $P_1(x,y)$. Inspired by traitor tracing, we consider a probabilistic surjection where $\mathbb{P}(\mathsf{r}(t)=1) = \theta_t$. The vector $\boldsymbol{\theta} \in [0,1]^{n+1}$ parametrizes the surjection. Denote by $\nabla_{\boldsymbol{\theta}}\mathsf{V}(t)$ the derivative w.r.t. θ_t. After some lengthy calculus:

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}}\mathsf{V}(t) &= n^{-1}K_1(p,\boldsymbol{\theta})(t - nK_2(p,\boldsymbol{\theta})), \qquad (22)\\
K_1(p,\boldsymbol{\theta}) &= \mathbb{P}(T=t)\Delta,\\
K_2(p,\boldsymbol{\theta}) &= \frac{\mathsf{h}'(P_1(0,1)) - \mathsf{h}'(\mathbb{P}(Y=1))}{\Delta},\\
\Delta &= \mathsf{h}'(P_1(0,1)) - \mathsf{h}'(P_1(1,1)).
\end{aligned}$$

It is not possible to cancel the gradient $\nabla_{\boldsymbol{\theta}}\mathsf{V}$. The optimal $\boldsymbol{\theta}$ thus lies on the boundary of the hypercube $[0,1]^{n+1}$. This makes the surjection deterministic. Assuming $\mathbb{P}(Y=1|X=0) < \mathbb{P}(Y=1|X=1)$, then $0 < K_1(p,\boldsymbol{\theta})$ and $0 < K_2(p,\boldsymbol{\theta}) \leq 1$ because $\mathsf{h}'(\cdot)$ is strictly decreasing. This makes $\nabla_{\boldsymbol{\theta}}\mathsf{V}(0) < 0$ and θ_0 must be set to the lowest possible value, *i.e.* $\theta_0 = 0$, to increase V at most. This is indeed the case for any θ_t with $t < K_2(p,\boldsymbol{\theta})$. In the same way, $\theta_n = 1$ and so is θ_t if $t > K_2(p,\boldsymbol{\theta})$. Yet, for a given $\boldsymbol{\theta}$, $K_2(p,\boldsymbol{\theta})$ ranges from 0 to 1 as p increases from 0 to 1. Therefore, $\boldsymbol{\theta} = (0,\ldots,0,1,\ldots,1)$ is optimal only over an interval of p.

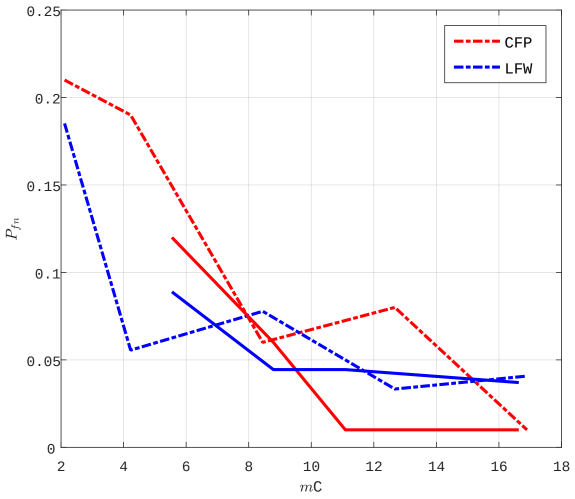

Fig. 4. Verification performance $P_{\mathsf{fn}}@P_{\mathsf{fp}} = 0.05$ *vs.* $m\mathsf{C}$, for $n = 16$. This quantity is reduced by decreasing m (dashed lines) or by decreasing C thanks to a surjection (solid lines).

For n odd and $p = 1/2$, $\theta_t = 0$ if $t \leq (n+1)/2$, and 1 (*i.e.* majority vote) otherwise is optimal because $K_2(1/2,\boldsymbol{\theta}) = 1/2$ ($\mathbb{P}(Y=1) = 1/2$ and $P_1(0,1) = 1 - P_1(1,1)$).

The 'All-1' surjection: $\boldsymbol{\theta} = (0,1,\ldots,1)$ makes $P_1(1,1) = 1$ so that $\nabla_{\boldsymbol{\theta}}\mathsf{V}(t) = +\infty$ if $t > 0$ and < 0 for $t = 0$.

B. Impact of the channel

Suppose that η is a parameter of the channel $W(\cdot|\cdot)$. Then

$$\frac{\partial\mathsf{V}}{\partial\eta} = \sum_{q,y} \frac{\partial P_1(q,y)}{\partial\eta} \log \frac{P_1(q,y)}{P_0(q,y)}, \tag{23}$$

because $\sum_{q,y} \frac{\partial P_1(q,y)}{\partial\eta} = \frac{\partial \sum_{q,y} P_1(q,y)}{\partial\eta} = 0$ and $\sum_{q,y} \frac{P_1(q,y)}{P_0(q,y)} \frac{\partial P_0(q,y)}{\partial\eta} = \sum_q \frac{\partial \mathbb{P}(Q=q)}{\partial\eta} = 0$.

Suppose now that $\eta = \eta_0 := W(q|0), \forall q \in \mathcal{X}\backslash 0$. Then,

$$\frac{\partial P_1(q,y)}{\partial\eta_0} = \mathbb{P}(X=0, Y=y)\,\forall q \in \mathcal{X}\backslash\{0\}. \tag{24}$$

Taking (23) around the noiseless channel where $\eta_0 = 0$ and $\mathbb{P}(X=0,Y=y) = P_1(0,y)$ because $Q = X$:

$$\left.\frac{\partial\mathsf{V}}{\partial\eta_0}\right|_{\eta_0=0} = \sum_{y,x\neq 0} P_1(0,y) \log \frac{P_1(x,y)}{P_0(x,y)} + \ldots \tag{25}$$

We only express the first terms to outline that if $P_1(x,y) = 0$ while $P_1(0,y)$ and hence $P_0(x,y)$ are not null, then this derivative goes to $-\infty$. A small deviation from the noiseless case with $\eta_0 \neq 0$ has a major detrimental impact on V. That situation happens for sure when working with type, *i.e.* $Y = T$: Consider the null type t_0 obtained when $X_1 = \ldots = X_n = 0$: $P_1(0,t_0) > 0$ while $P_1(x,t_0) = 0$, $\forall x \neq 0$.

One can prove that the surjection can mitigate this effect if $\exists t \neq t_0 : \mathsf{r}(t) = \mathsf{r}(t_0)$ and $P_1(0,t) > 0$. This happens with the majority vote of the dense setup, but unfortunately, not with of the 'All-1' surjection in the sparse setup.

References

[1] M. Gheisari, T. Furon, L. Amsaleg, B. Razeghi, and S. Voloshynovskiy, "Aggregation and embedding for group membership verification," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.

[2] M. Gheisari, T. Furon, and L. Amsaleg, "Privacy preserving group membership verification and identification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[3] B. Razeghi, S. Voloshynovskiy, D. Kostadinov, and O. Taran, "Privacy preserving identification using sparse approximation with ambiguization," in *Proceedings of the IEEE International Workshop on Information Forensics and Security*, 2017.

[4] B. Razeghi and S. Voloshynovskiy, "Privacy-preserving outsourced media search using secure sparse ternary codes," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018.

[5] C. E. Shannon, "Probability of error for optimal codes in a gaussian channel," *Bell System Tech. J.*, vol. 38, pp. 611–656, 1959.

[6] T. Laarhoven, "Search problems in cryptography from fingerprinting to lattice sieving," Ph.D. dissertation, Eindhoven University of Technology, 2015.

[7] J. Boersma, "Solution to problem 87-6* : The entropy of a poisson distribution," *SIAM Review*, vol. 30, no. 2, pp. 314–317, 1988.

[8] A. Andoni, P. Indyk, T. Laarhoven, I. P. Razenshteyn, and L. Schmidt, "Practical and optimal LSH for angular distance," *NIPS*, 2015. [Online]. Available: http://arxiv.org/abs/1509.02897

[9] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database forstudying face recognition in unconstrained environments,” in *Workshop on faces in’Real-Life’Images: detection, alignment, and recognition*, 2008.

[10] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, and D. W. Jacobs, “Frontal to profile face verification in the wild,” in *Proceeding of the IEEE Winter Conference on Applications of Computer Vision*, 2016.

[11] C. E. Thomaz and G. A. Giraldi, “A new ranking method for principal components analysis and its application to face image analysis,” *Image and Vision Computing*, vol. 28, no. 6, pp. 902–913, 2010.

[12] O. M. Parkhi, A. Vedaldi, A. Zisserman *et al.*, “Deep face recognition.” in *Proceedings of the British Machine Vision Conference*, 2015.

[13] M. E. Tipping and C. M. Bishop, “Probabilistic principal component analysis,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.