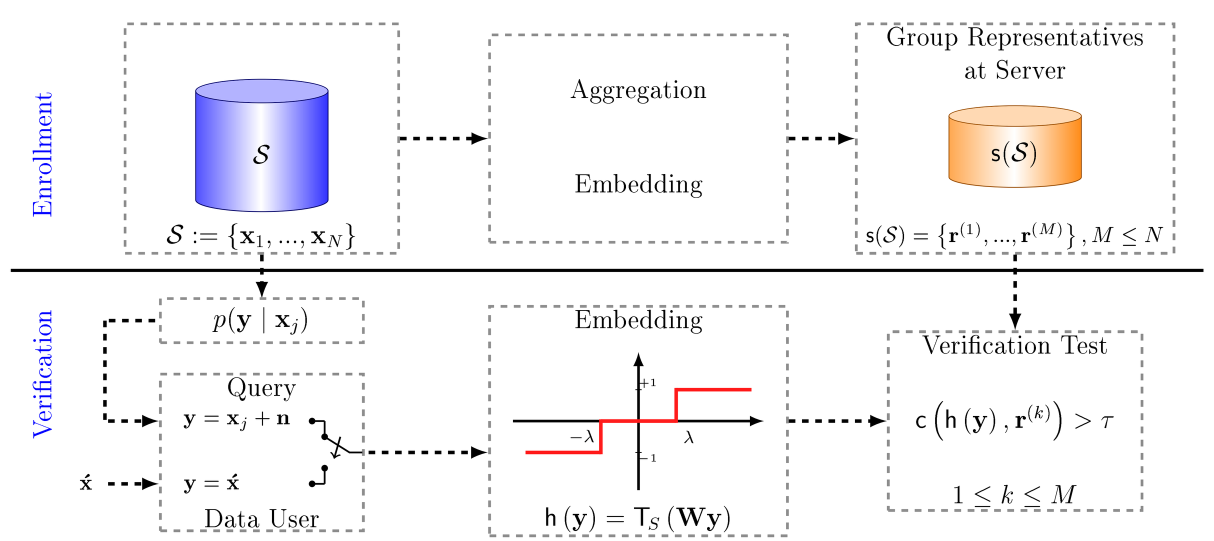
Aggregation and Embedding for Group Membership Verification

Marzieh Gheisari†, Teddy Furon†, Laurent Amsaleg†, Behrooz Razeghi⋆, Slava Voloshynovskiy⋆
† Univ Rennes, Inria, CNRS, IRISA, France, ⋆University of Geneva, Switzerland

Introduction

- **Group Membership Verification**
  - Verify an item is part of a group
  - Without identifying this item (privacy)
- **Feature extraction**
  - An item = device, person, object
  - We extract a *signature* $\mathbf{x} \in \mathbb{R}^d$ per item (PUF, biometric traits, descriptor,...)
- **Protocols**
  - Enrollment: A data structure memorizes a group of signatures, stored by server
  - Verification: The data structure is queried by a client signature $\mathbf{y} \in \mathbb{R}^d$
- **Security**
  - The data structure protects enrolled signatures against honest but curious server
  - Verification proceeds with privacy, not disclosing identity

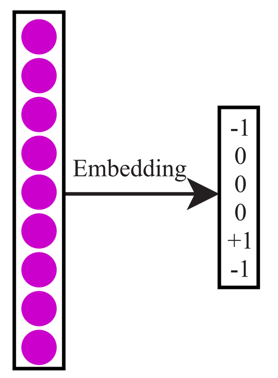
Problem Formulation



- Testing hypothesis about query $\mathbf{y}$

$\mathcal{H}_1$ $\mathbf{y} = \mathbf{x}_j + \mathbf{n}$ with $\mathbf{x}_j$ enrolled signature
$\mathcal{H}_0$ $\mathbf{y}$ not related to any signature in the group

- Privacy enabling embedding [1]
  - $\mathsf{h} : \mathbb{R}^d \to \{-1, 0, +1\}^l$
  - Properties
    - Sparsity: $\|\mathsf{h}(\mathbf{x})\|_0$ small
    - Inaccurate reconstruction

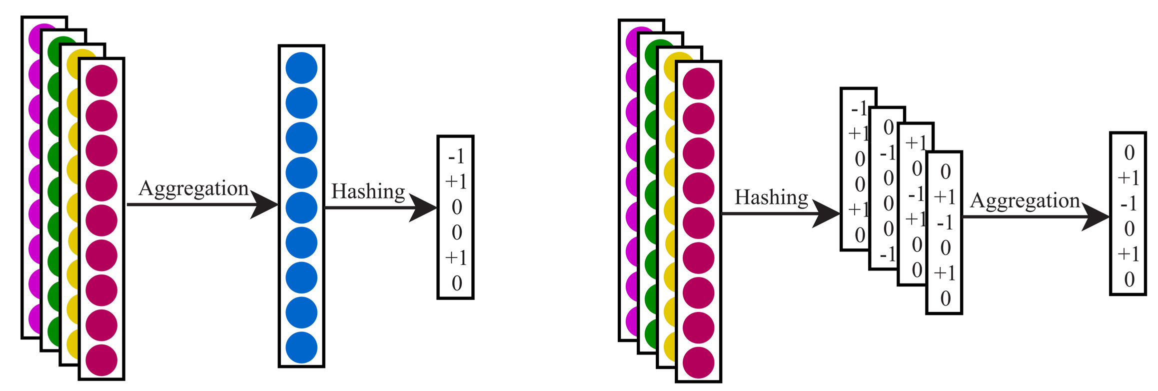


References

[1] *"Privacy preserving identification using sparse approximation with ambiguization,"* B. Razeghi, S. Voloshynovskiy, D. Kostadinov, and O. Taran, in Proc. of IEEE WIFS, 2017.
[2] *"Memory vectors for similarity search in high-dimensional spaces,"* A. Iscen, T. Furon, V. Gripon, M. Rabbat, and H. Jégou, IEEE Trans. on Big Data, 2017.

Aggregation Strategies

- **HoA**: Aggregate first, then embed $\mathsf{s} = \mathsf{h} \circ \mathsf{a}$
  - Sum: $\mathsf{a}(\mathcal{S}) = \sum_{\mathbf{x} \in \mathcal{S}} \mathbf{x} = \mathbf{G}\mathbf{1}_N$
  - Pseudo-inverse [2]: $\mathsf{a}(\mathcal{S}) = (\mathbf{G}^{\dagger})^{\top}\mathbf{1}_N$
- **AoH**: Embed first, then aggregate $\mathsf{s} = \mathsf{a} \circ \mathsf{h}$
  - Sum: $\mathbf{r} = \mathsf{sign}(\sum_{\mathbf{x} \in \mathcal{S}} \mathsf{h}(\mathbf{x}))$
  - Sign-pooling: $r_i = \arg\max_s |\{\mathbf{x} \in \mathcal{S} | \mathsf{h}(\mathbf{x})_i = s\}|$

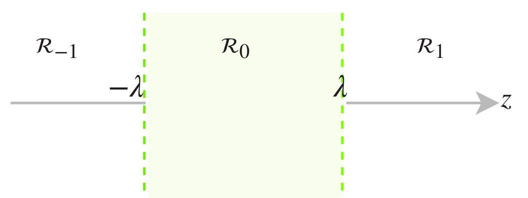


Metrics

- Verification Performances
  - Usual hypotheses testing metrics A.U.C. or $p_{\mathsf{fn}}(\tau)$ for τ s.t. $p_{\mathsf{fp}}(\tau) = \epsilon$
- Security and Privacy
  - Measure ability to reconstruct signatures from query or group representation

Reconstruction of the Query

- Assumptions
  - $\mathbf{y} \sim \mathcal{N}(\mathbf{0}_d, \sigma_y^2 \mathbf{I}_d)$
  - $\mathbf{W}$ orthogonal matrix known by the attacker
- Information leakage
  - Consider $\mathbf{z} = \mathbf{W}\mathbf{y}$
  - Observing the i-th symbol of $\mathsf{h}(\mathbf{y})$ equals s reveals that $z_i \in \mathcal{R}_s$



- Optimal reconstruction is component-wise

$$\hat{z}_i(s) := \mathbb{E}(Z|\mathcal{R}_s) = \begin{cases} 0 & \text{if } s = 0 \\ \frac{s\sigma_y}{p_1\sqrt{2\pi}} e^{-\frac{\lambda^2}{2\sigma_y^2}} & \text{otherwise} \end{cases}$$

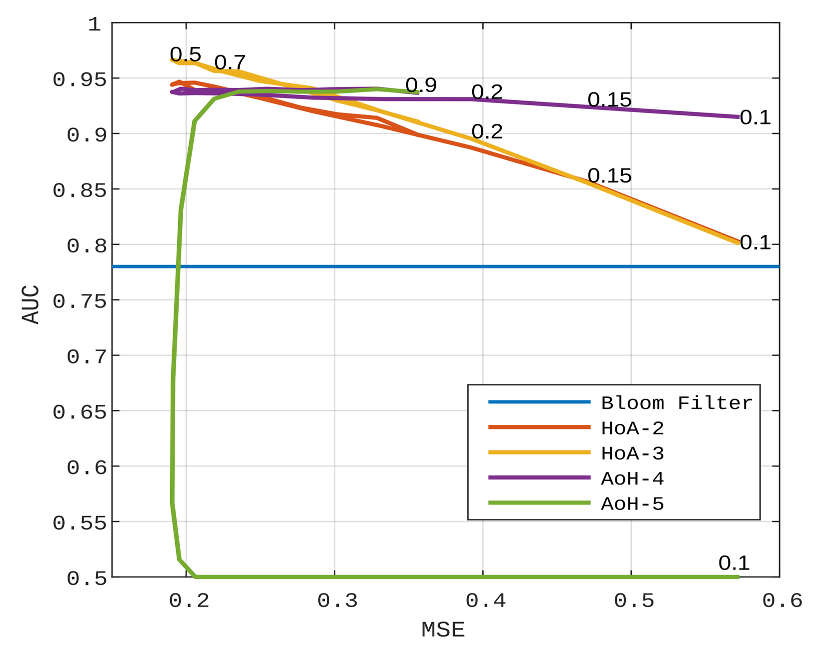
$$\frac{\mathsf{MSE}_q}{\sigma_y^2} = \mathsf{MSE}(\lambda) = 1 - \frac{1}{\pi\Phi(-\lambda/\sigma_x)} e^{-\frac{\lambda^2}{\sigma_y^2}}$$

Reconstruction of Enrolled Signatures

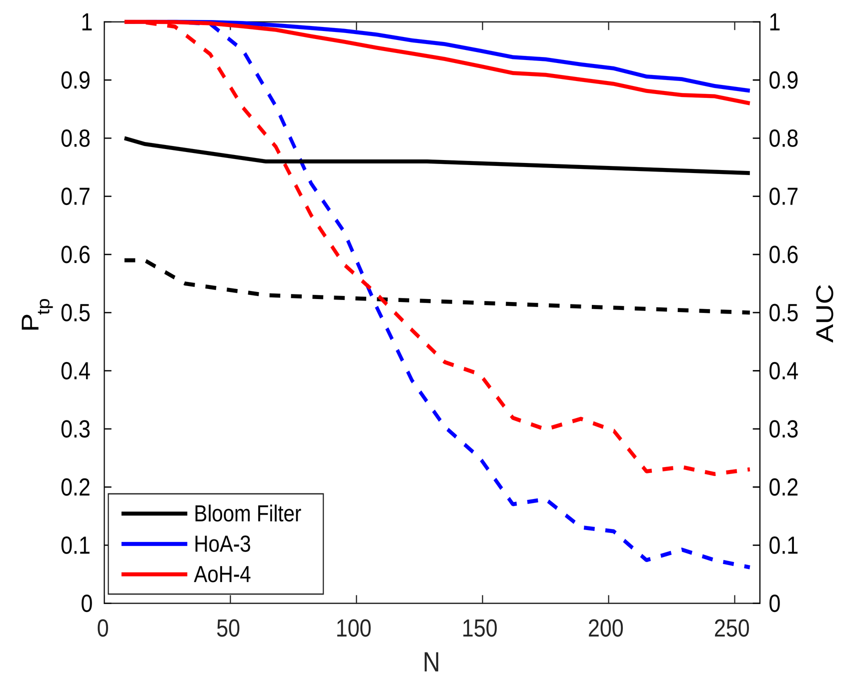
- Reconstructing $\hat{\mathbf{x}}$ from group representation $\mathbf{r}$: Unique reconstruction for all group signatures $\mathsf{MSE}_e = (dN)^{-1}\sum_{j=1}^{N} \mathbb{E}(\|\mathbf{X}_j - \hat{\mathbf{X}}\|^2)$
- Assume $\hat{\mathbf{x}} = \kappa\mathbf{u}$
  - The best choice is $\kappa = \|\mathbf{u}\|^{-2}\mathbf{u}^{\top}\mathbf{m}$ and $\mathbf{u} \propto \mathbf{m}$, with $\mathbf{m} := N^{-1}\sum_{j=1}^{N} \mathbf{x}_j$
  - Reconstructing $\hat{\mathbf{m}}$ is only possible for HoA - Sum
  - Lower bound: $\mathsf{MSE}_e \geq \sigma_x^2 \left(1 - \frac{1}{N}(1 - \mathsf{MSE}(\lambda))\right)$

Verification Performances

- Baseline
  - A Bloom filter optimally tuned for N and p_{fp} with $\ell_B = \lceil N|\log p_{\mathsf{fp}}|\log(2)^{-2}\rceil$ bits
- Verification performance vs. $\mathsf{MSE}_q/\sigma_y^2$ ($N = 128$, $d = 1024$, $\sigma_n^2 = 0.01$ and $S/d \in (0.1, 0.9)$)



- Verification performance vs. N (Solid and dashed lines correspond to AUC and $p_{\mathsf{tp}}@p_{\mathsf{fp}} = 10^{-2}$)

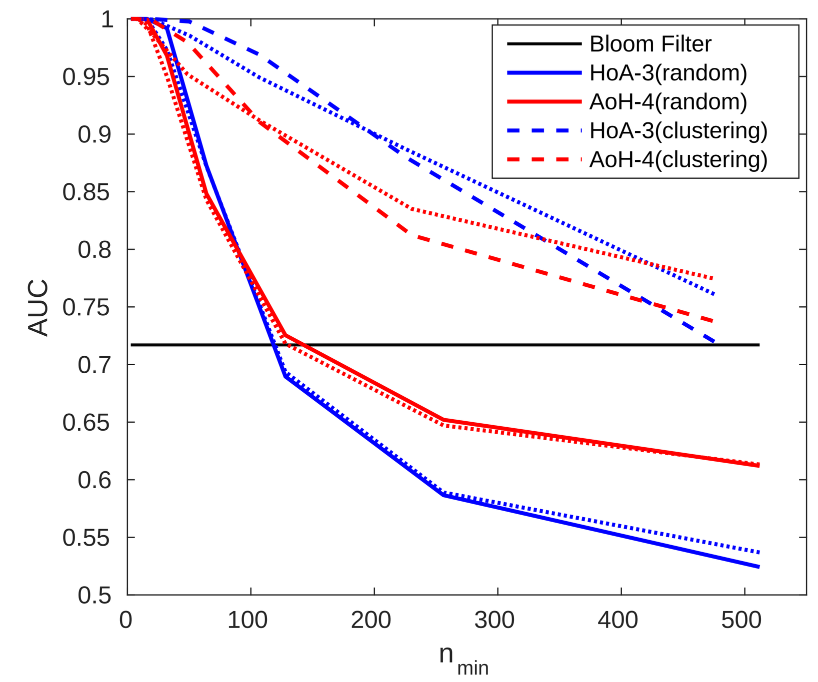


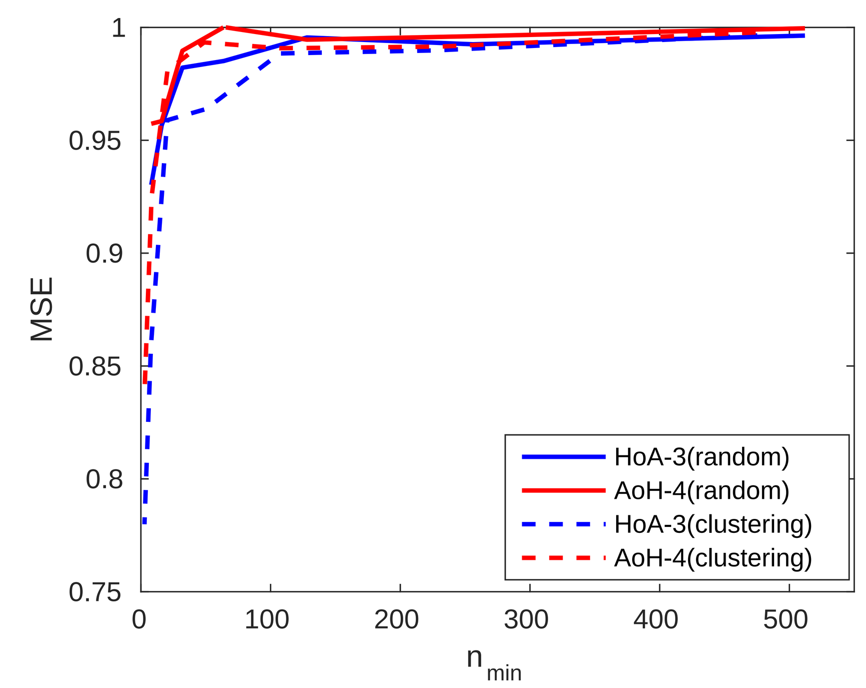
Several Groups

- Enroll N signatures into $M > 1$ groups
  - Random assignment (M groups of size $n = N/M$) or by Clustering (k-means algorithm)

$$n_{\min} = \min_{1 \leq k \leq M}(n_k)$$

- Setup: $N = 4096$, $d = 1024$, $\sigma_n^2 = 10^{-2}$, $S/d = 0.6$ for HoA-3, and $S/d = 0.85$





- Conclusion: Clustering boosts verification performances while not endangering the security