Privacy-Preserving Near Neighbor Search via Sparse Coding with Ambiguation

Behrooz Razeghi*, Sohrab Ferdowsi†, Dimche Kostadinov‡, Flavio. P. Calmon§, Slava Voloshynovskiy*
*University of Geneva
†HES-SO Geneva
‡University of Zurich
§Harvard University

***Abstract*—In this paper, we propose a framework for privacy-preserving approximate near neighbor search via stochastic sparsifying encoding. The core of the framework relies on sparse coding with ambiguation (SCA) mechanism that introduces the notion of inherent shared secrecy based on the support intersection of sparse codes. This approach is 'fairness-aware', in the sense that any point in the neighborhood has an equiprobable chance to be chosen. Our approach can be applied to raw data, latent representation of autoencoders, and aggregated local descriptors. The proposed method is tested on both synthetic i.i.d data and real large-scale image databases.**

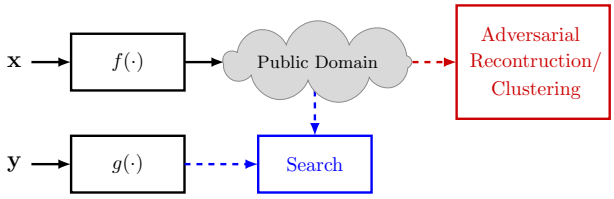

Fig. 1: The general block diagram of our framework.

I. Introduction

Many modern signal processing, machine learning and data mining applications, such as biometric authentication/identification, pattern recognition, speech processing and recommender systems, require near neighbor search of a query with respect to a given dataset, and a distance measure. Many search services are outsourced to third parties (service providers) who possess powerful storage, communications and computing facilities. The major challenge is to satisfy privacy constraints of the owner's data and the clients' interests, while still being capable of performing the fast search service in multi-billion entry datasets.

Let $(\mathcal{S}, d_{\mathcal{S}})$ be a metric space. Given a set $\mathcal{X} \subseteq \mathcal{S}$ of M points, a parameter r, and a query point $\mathbf{y} \in \mathcal{S}$, the goal of the exact near neighbor (NN) problem is to find a point $\mathbf{x} \in \mathcal{X}$ such that $d_{\mathcal{S}}(\mathbf{x}, \mathbf{y}) \leq r$, if such a point exists. In the approximate variant of this problem (ANN), given $c > 1$, the problem is relaxed to find a point $x \in \mathcal{X}$ such that $d_{\mathcal{S}}(\mathbf{x}, \mathbf{y}) \leq cr$. These problems can be generalized to k-NN and k-ANN setting. In this context, we assume $\mathcal{S}$ to be the N-dimensional Euclidean space, i.e., $\mathcal{S} = \mathbb{R}^N$, and the distance given by an ℓ_2-norm, $d_{\mathcal{S}}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2$. The NN search based on naïve solution, i.e., the linear scan, is the bottleneck of the system in large scale high-dimensional data sets [1]. Alternatively, approximate near neigbor (ANN) search, is more efficient in terms of query time and space complexity [2]–[4]. Perhaps the most popular solution to ANN problem is via hashing, where the aim is to *transform* the data points to a lower dimensional space, then perform similarity search in the lower dimensional representation. The two main research directions are (1) Locality Sensitive Hashing (LSH): indexing points using a hash table with the property that similar (closer) data points have a higher probability of collision than dissimilar (far) points [1]–[7]; (2) learning to hash: performing NN similarity search in a low dimensional space with a lower search complexity [1], [8]–[11]. The objective in the latter methodology is to preserve *semantic* or *distance similarity* between the original space and transformed space. Subsequent research showed that quantization-based solutions are preferred in terms of query time, space cost and search accuracy [1].

Concretely, this work brings the following contributions: (1) We consider the methodology of learning to hash for privacy-preserving proximity search which entails minimum information loss for authorized users. The authorized parties can *purify* the ambiguation noise using the *shared secrecy* based on the support intersection of sparse codes [12].
(2) We adopt a notion of fairness addressed in [7], but in a privacy-preserving setup. Our notion of fairness differs from machine learning algorithms where the goal is to handle the bias introduced at the training phase. We consider the bias in the *stored data* and *querying response*. By doing this, any point in the neigborhood has an equal chance to be chosen. Moreover, in some cases, it may suffice to return any of the points in the near neighborhood, rather than the computationally expensive nearest one. The equiprobable nearby scheme can also be utilized in privacy protection mechanisms. That is, instead of reporting the nearest-neighbor, which leaks more information, the service provider just sends back a random or a typical data point close to the query point.

In comparison to [13], [14], our work has the following fundamental differences: a) In [13]–[16], they utilized a dimensionality reduction transform with random entries, while our sparsifying transform may keep, extend, or reduce the dimension of the original data. Moreover, our transform is

learned using the sparsifying transform problem to ensure an optimal sparse representational that is information preserving in general, whereas the transform in [13]–[15] might preserve the distances only under certain conditions of the Johnson-Lindenstrauss Lemma. b) In [13], [14], [16], the codes are dense and binary, whereas in our method the codes are sparse (and possibly ternary), which form a basis of our ambiguation framework. Last but not the least, the embedding based on universal quantization scheme [13] has information leakage in terms of clustering, i.e., the curious server still can perform clustering on data points. Moreover, we impose no restrictions on the input data, i.e., we assume that as an input we might have raw data, extracted features using any known hand crafted methods, aggregated local descriptors based on BoW, FV, VLAD [17]–[19], etc., or from the last layers of deep nets [20], or the latent space of auto-encoders [21]. We apply our model on the latent representation of a designed network in [12].

Throughout this paper, superscript $(\cdot)^T$ stands for the transpose. Vectors and matrices are denoted by boldface lower-case ($\mathbf{x}$) and upper-case ($\mathbf{X}$) letters, respectively. We consider the same notation for a random vector $\mathbf{x}$ and its realization. The difference should be clear from the context. x_i denotes the i-th entry of vector $\mathbf{x}$. For a matrix $\mathbf{X}$, $\mathbf{x}(j)$ denotes the j-th column of $\mathbf{X}$. We use the notation $[N]$ for the set $\{1, 2, ..., N\}$.

II. Preliminaries

A. Problem Setup

Consider a three-party data release scenario involving (a) a data owner, (b) data users, and (c) a service provider (server). The data owner possesses database $\mathbf{X} = [\mathbf{x}(1), \cdots, \mathbf{x}(M)]$ consisting of M data points $\mathbf{x}(m) \in \mathbb{R}^N$, $m \in [M]$. The database is used to offer some utility for the *authorized* data users. The data users seek some utility from the data owner based on their query $\mathbf{y}$. The server provides a predetermined service to the data users on behalf of the data owner. We assume that both the server and data users are honest-but-curious, which we consider them as an adversary. The service provider may try to infer some information about the original data collection $\mathcal{X}$ from the disclosed public storage and/or the querying data sent to the server. For instance, the server may estimate the original data from the disclosed representations and query, or may establish links between the closet entries in database. The data users may try to infer some information about the public representations and/or the original data via multiple varied queries to guess the data manifold by inspecting the returned responses. A general diagram of our framework is depicted in Fig. 1.

Therefore, we study the problem of disclosing database $\mathbf{X}$ to a third-party (public storage) in order to drive some utility, in terms of near neighbor search, for the authorized data users based on the public representations while, at the same time, protect the privacy of the data owner (against the honest-but-curious server and data users) and data users (against the honest-but-curious server).

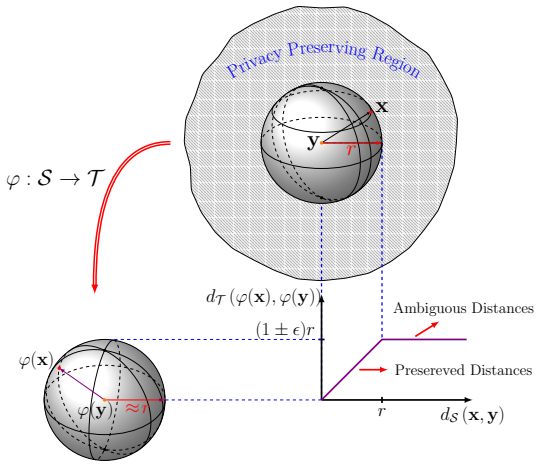

Fig. 2: Visualization of desired property of mapping scheme in privacy-preserving near neighbor search setup.

B. Fair Near Neighbor

Let $(\mathcal{S}, d_{\mathcal{S}})$ be a metric space and let $\mathcal{X} \subseteq \mathcal{S}$ be a set of M data points. Let $B_{\mathcal{S}}(\mathbf{c}, r) = \{\mathbf{x} \in \mathcal{S} \mid d_{\mathcal{S}}(\mathbf{c}, \mathbf{x}) \leq r\}$ be the closed ball of radius $r > 0$ around a point $\mathbf{c} \in \mathcal{S}$. Let $N(\mathbf{c}, r) = B_{\mathcal{S}}(\mathbf{c}, r) \cap \mathcal{X}$ be the r-neighborhood of $\mathbf{c}$ in $\mathcal{X}$, with the size $|N(\mathbf{c}, r)|$.

Definition 1. Fair Near Neighbor (FNN) [7]. Given a data set $\mathcal{X} \subseteq \mathcal{S}$ of M data points, a parameter $r > 0$, and query point $\mathbf{y}$, the goal is to find a data point $\mathbf{x} \in N(\mathbf{y}, r)$ with probability μ, where $1/(|N(\mathbf{y}, r)|(1+\epsilon)) \leq \mu \leq (1+\epsilon)/|N(\mathbf{y}, r)|$, i.e., μ is an approximately uniform probability distribution.

C. Fair Privacy-Preserved Approximate Near Neigbor

Let $(\mathcal{T}, d_{\mathcal{T}})$ be a metric space where $d_{\mathcal{T}}(\mathbf{a}, \mathbf{b}), \forall \mathbf{a}, \mathbf{b} \in \mathcal{T}$ is defined on $\text{supp}(\mathbf{a}) = \{l \in [L] : a_l \neq 0\}$. Let $\mathcal{P} \subseteq \mathcal{T}$ be a set of M data points. Let $B_{\mathcal{T}}(g(\mathbf{y}), r) = \{f(\mathbf{x}) \in \mathcal{T} \mid d_{\mathcal{T}}(g(\mathbf{y}), f(\mathbf{x})) \leq r\}$. Now, we define the *Fair Privacy-preserved Approximate Near Neighbor* method as follows:

Definition 2. Fair Privacy-preserved Approximate Near-Neighbor (FPANN). Given a data set $\mathcal{X} \subseteq \mathcal{S}$ of M data points, a parameter $r > 0$, the goal is to design a randomized privacy-preserving data release mechanism $f : \mathcal{X} \to \mathcal{P}$ and (randomized) query processing $g : \mathcal{S} \to \mathcal{T}$ such that for a given *authorized* query $\mathbf{y}$ one can report a point $f(\mathbf{x}) \in N_{\mathcal{T}}(g(\mathbf{y}), r)$ with probability μ_p, where $N_{\mathcal{T}}(g(\mathbf{y}), r) = B_{\mathcal{T}}(g(\mathbf{y}), r) \cap \mathcal{P}$ be the approximate r-neighbor of $g(\mathbf{y})$ in $\mathcal{P}$, and $1/(|N(g(\mathbf{y}), r)|(1+\epsilon)) \leq \mu_p \leq (1+\epsilon)/|N(g(\mathbf{y}), r)|$.

Definition 3. (β, γ)-recoverable privacy mechanism. For $0 \leq \gamma \leq 1$ and given authorized query $\mathbf{y}^{\text{auth}}$, unauthorized query $\mathbf{y}^{\text{unauth}}$ and $\beta > 0$, a privacy-preserving data release mechanism $f : \mathcal{X} \to \mathcal{P}$ is (β, γ)-recoverable if:

$$
\begin{aligned}
(i): \quad P_e^{\text{auth}} &= \Pr\left[\mathbb{E}\left[d(\mathbf{x}, \hat{\mathbf{x}})\right] \leq \beta \mid g\left(\mathbf{y}^{\text{auth}}\right)\right] < \gamma, \\
(ii): P_e^{\text{unauth}} &= \Pr\left[\mathbb{E}\left[d(\mathbf{x}, \hat{\mathbf{x}})\right] \leq \beta \mid g\left(\mathbf{y}^{\text{unauth}}\right)\right] \geq \gamma,
\end{aligned}
$$

where $g(\cdot)$ is the data user's query function to service provider.

III. Proposed Framework

A. Framework Overview

Our framework is composed of the following steps:
1) *Preparation at Owner Side:* The owner generates the sparse codewords from the data that s/he owns using the *trained sparsifying transform*. Next, he shares the privacy-protected sparse codebook with the service provider (server). Following Kerckchoffs's Principle in cryptography, the data owner makes the learned sparsifying transform publicly available.
2) *Indexing at Server Side:* The server indexes the received sparse codes in a database.
3) *Querying at Data User Side:* The data user generates a sparse representation from his query data using the shared transform. Then, the client sends a function of his sparse representation to the server.
4) *Near Neighbor Search at Server Side:* Given the requested probe, the server runs a near neighbor search to find the stored sparse codes that are most similar (close) to the probe. Finally, based on the pre-determined service to the data users, the server sends back an answer to the data user.

Next, we describe in more detail the fundamental elements of our mechanism.

B. Sparse Data Representation

The goal of sparsification is to obtain an information-preserving sparse representation of the original data. Our sparsifying transform consists of a linear mapper followed by an element-wise nonlinearity. We consider a joint learning problem to obtain the sparsifying transform $\mathbf{W} \in \mathbb{R}^{L\times N}$ as well as the sparse codebook $\mathbf{A} \in \mathbb{R}^{L\times M}$ that can be formulated as:

$$\left(\hat{\mathbf{W}},\hat{\mathbf{A}}\right)=\arg\min_{(\mathbf{W},\mathbf{A})} \|\mathbf{W}\mathbf{X}-\mathbf{A}\|_F^2+\beta_1\Omega_1(\mathbf{W})+\beta_2\Omega_2(\mathbf{A})\,, \quad (1)$$

where $\beta_1 \geq 0$ and $\beta_2 \geq 0$ are regularization parameters, $\Omega_1(\mathbf{W}) = (\frac{1}{\beta_{1,1}}\|\mathbf{W}\|_F^2 + \frac{1}{\beta_{1,2}}\|\mathbf{W}\mathbf{W}^T - \mathbf{I}\|_F^2 - \frac{1}{\beta_{1,3}}\log|\det \mathbf{W}^T\mathbf{W}|)$ penalizes the information loss in order to avoid trivial solutions, and $\Omega_2(\mathbf{A})$ is the sparsity constraint on the compressed codebook $\mathbf{A}$ [22]. The term $\|\mathbf{W}\mathbf{X}-\mathbf{A}\|_F^2$ is a sparsification error, which represents the deviation of the transformed data from the exact sparse representation in the transformed domain [1] Our algorithm for solving (1) alternates between a ℓ_0-"norm"-based *sparse coding step*, and a non-convex *transform update step* [26]. Therefore, one can write the closed-form formulation of the *encoder* as [26], [27]:

$$\mathbf{a}(m) = \varphi\left(\mathbf{x}(m)\right) = \psi_\lambda\left(\mathbf{W}\mathbf{x}(m)\right), \forall m \in [M]\,, \quad (2)$$

where $\psi_\lambda(\mathbf{f}) = 1_{|f_l|\geq\lambda}\mathbf{f}, \forall l \in [L], \lambda \geq 0$ and $\mathbf{a}(m)$ is S_x-sparse, i.e., $\|\mathbf{a}(m)\| \approx S_x, \forall m \in [M]$. The decoder (reconstruct mapper) $\mathbf{R} \in \mathbb{R}^{N\times L}$ can be formulated as follows:

$$\begin{aligned} \min_{\mathbf{R}} \quad & \|\mathbf{R}\mathbf{A}-\mathbf{X}\|_F^2 +\beta_R\|\mathbf{R}-\left(\mathbf{W}^T\mathbf{W}+\beta\mathbf{I}\right)^{-1}\mathbf{W}^T\|_F^2, \\ \text{st}: \quad & \mathbf{R}^T\mathbf{R} = \mathbf{I}, \end{aligned} \quad (3)$$

[1] We refer the reader to [23]–[25] for applications in group membership verification.

where $\mathbf{A} \in \mathbb{R}^{L\times M}$ is sparse codebook, $\mathbf{X} \in \mathbb{R}^{M\times N}$ is original data points and $\mathbf{W} \in \mathbb{R}^{L\times N}$ is encoder transform. Since $\mathbf{R}$ has orthonormal columns, we have $\|\mathbf{R}\mathbf{A}-\mathbf{X}\|_F^2 = \mathrm{tr}\left[\mathbf{X}^T\mathbf{X}-2\mathbf{X}^T\mathbf{R}\mathbf{A}+\mathbf{A}^T\mathbf{A}\right], \|\mathbf{R}-\left(\mathbf{W}^T\mathbf{W}+\beta\mathbf{I}\right)^{-1}\mathbf{W}^T\|_F^2 = \mathrm{tr}\left[\mathbf{I}-2\mathbf{C}^T\mathbf{R}+\mathbf{C}^T\mathbf{C}\right]$, where $\mathbf{C} = \left(\mathbf{W}^T\mathbf{W}+\beta\mathbf{I}\right)^{-1}\mathbf{W}^T$. Consequently (3) is equivalent to the problem of maximizing $\mathrm{tr}\left[\mathbf{X}^T\mathbf{R}\mathbf{A}\right]+\beta_R\mathrm{tr}\left[\mathbf{C}^T\mathbf{R}\right] = \mathrm{tr}\left[\left(\mathbf{A}\mathbf{X}^T+\beta_R\mathbf{C}^T\right)\mathbf{R}\right]$. Considering the Singular Value Decomposition $\mathbf{A}\mathbf{X}^T+\beta_R\mathbf{C}^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, this formulation reduces to $\mathrm{tr}\left[\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{R}\right] = \mathrm{tr}\left[\mathbf{\Sigma}\mathbf{Z}\right] = \sum_i z_{ii}\Sigma_{ii} \leq \sum_i \Sigma_{ii}$, where $\mathbf{Z} = \mathbf{V}^T\mathbf{R}\mathbf{U}$. Note that the last inequality holds because $\mathbf{Z}$ is an orthonormal matrix, and $\sum_j z_{ij}^2 = 1$, $z_{ii} \leq 1$. Therefore, the maximum can be achieved if $\mathbf{Z} = \mathbf{I}$, i.e., closed form solution is $\mathbf{R} = \mathbf{U}\mathbf{V}^T$, where $\mathbf{A}\mathbf{X}^T+\beta_R\left(\left(\mathbf{W}^T\mathbf{W}+\beta\mathbf{I}\right)^{-1}\mathbf{W}^T\right)^T = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

C. Ambiguation Mechanism

The idea of ambiguation is to add (pseudo) random noise to the orthogonal complement, i.e., non-informative components of the sparse code. The integration of 'sparse lossy coding' with 'ambiguation' introduces a generalized randomization technique, namely Sparse Coding with Ambiguation (SCA) [27]. The SCA provides an information-theoretically and computationally private mechanism. The information-theoretical privacy guarantee originate from the lossy compression induced at the sparsification stage, and the computational privacy guarantee originate from ambiguation stage. The curious server faces a combinatorial complexity budget requirement to guess the informative components. The ambiguation noise is required to have the same distribution as the sparse codes, to guarantee being indistinguishable from its statistical properties. We refer the reader to [27] for more details. The randomized privacy-preserving data release mechanism $f : \mathcal{X} \to \mathcal{P} \subseteq \mathcal{T}$ can be formulated as:

$$\mathbf{p}(m) = f(\mathbf{x}(m)) = \varphi(\mathbf{x}(m)) \oplus \mathbf{n}^p_{\text{supp}},\ \forall m \in [M]\,, \quad (4)$$

where $\|\mathbf{n}^p_{\text{supp}}\|_0 \approx S_p$.

Given a query point $\mathbf{y}$, the (randomized) query release mechanism $g : \mathcal{S} \to \mathcal{T}$ can be formulated as:

$$\mathbf{q} = g(\mathbf{y}) = \varphi\left(\mathbf{y}\right) \oplus \mathbf{n}^q_{\text{supp}}, \quad (5)$$

where $\|\mathbf{n}^q_{\text{supp}}\|_0 \leq S_q$, $0 \leq S_q \leq S_p$. If $S_q = 0$, the query is disclosed as in-the-clear sparse code without ambiguation noise. Let us consider two hypotheses for near neighbor search as follows. $\mathcal{H}_1$: The authorized query is related to one of the M data points in the database. For instance, it is a noisy version of one data point. $\mathcal{H}_0$: The unauthorized query is not related to any data point. For instance, it is synthetic query generated by an adversary.

D. Near Neighbor Search

The near neighbor search is performed in latent space $\mathcal{T}$. Given a data set $\mathcal{P}$ of M embedded disclosed representations $\{\mathbf{p}(m)\}, m \in [M]$, a parameter r, and embedded query point $\mathbf{q}$, the service provider performs approximate near neighbor search and report a point randomly and uniformly from $B_{\mathcal{T}}\left(\mathbf{q}, r\right) \cap \mathcal{P}$.

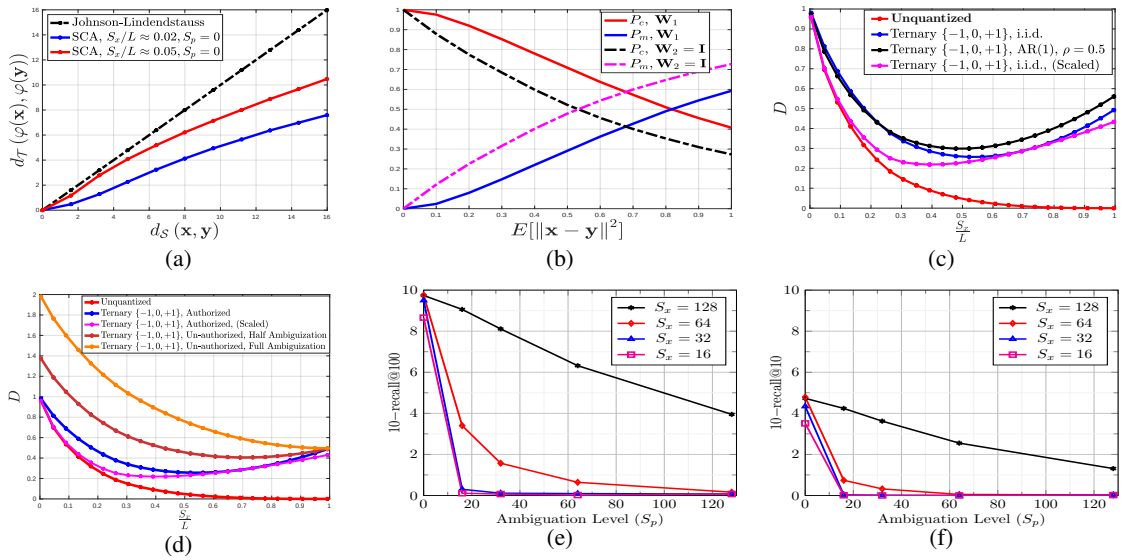

Fig. 3: a) local distance preserving; b) local robustness; (c) and (d): Comparison of distortion-sparsity behavior for c) authorized and d) unauthorized parties; (e) and (f) $R-$recall@T curves for a subset of 10K CelebA images of $3 \times 128 \times 128$.

IV. Discussion

We now discuss various properties of our method. One desired property of an embedding scheme in a privacy-preserving near neighbor search is to preserve distance information only up to a specified radius, while quickly flattening after this distance threshold. Therefore, from one hand, the information rate is spent in encoding local distances, and from the other hand, the curious server/data user cannot recover any distance information about signals that are far apart. Fig. 2 visualizes this local isometric mapping, where $d_{\mathcal{S}}(\cdot,\cdot)$ and $d_{\mathcal{T}}(\cdot,\cdot)$ denote the distance measure in original domain and transform domain, respectively.

Suppose $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{x}}\mathbf{I}_N)$ and $\mathbf{y}^{\text{auth}} = \mathbf{x} + \mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma_{\mathbf{z}}\mathbf{I}_N)$, and where $\mathbf{x} \in \mathbf{X}$. Fig.3a depicts the behaviour of our embedding for two sparsity levels and compare them with linear embeddings which preserve all distances equally. Let us define:

$$P_c = \frac{1}{M.S_x} \sum \Pr\{\text{supp}\,(\varphi(\mathbf{x})) = \text{supp}\,(\varphi(\mathbf{y}))\},$$
$$P_m = \frac{1}{M.S_x} \sum \Pr\{\text{supp}\,(\varphi(\mathbf{x})) \neq \text{supp}\,(\varphi(\mathbf{y}))\}.$$

Fig.3b illustrates the probability of correct support and missed support for the learned linear map $\mathbf{W}_1$ (problem (1)) and $\mathbf{W}_2 = \mathbf{I}$. The learned transform outperforms in local distances.

A. Reconstruction Leakage

A potential threat is that the adversary may try to reconstruct the original data points from the disclosed representations. To get insight into the SCA model, we firstly provide the results on a synthetic database establish its connection to classical Shannon rate-distortion theory. Next, we validate our model on real image databases. For the sake of completeness, we also bring the results provided in [11] on synthetic i.i.d data. Note that the sparsity level S_x controls the information encoding rate, or equivalently, the distinguishability of data points in the transform domain. The ambiguation level S_p controls the imposed randomness to the informative data.

Fig. 3c and Fig. 3d illustrate and compare distortion-sparsity behavior at *authorized* and *unauthorized* parties, respectively. Fig.3c depicts reconstruction fidelity for four cases: 1) unquantized sparsifying encoding (2), 2) Sparse Ternary Coding (STC) for independent and identically distributed (i.i.d.) data [27], 3) STC for i.i.d. data which re-scaled in original domain [11], and 4) STC for correlated data which are drawn from AR(1) model with the parameter $\rho = 0.5$. We used the same experimental setting as [11]. Fig.3d shows the reconstruction leakage at a curious server (or an adversary) who knows the encoder and its parameters, but has no knowledge about the correct indices to purify the ambiguated representations. The

		$S_p = 0$					$S_x = 16$				
		$S_x = 1$	$S_x = 2$	$S_x = 4$	$S_x = 8$	$S_x = 16$	$S_p = 0$	$S_p = 8$	$S_p = 16$	$S_p = 24$	$S_p = 32$
MNIST	MSE	0.7331	0.6351	0.5101	0.3758	0.2726	0.2726	0.4030	0.4963	0.5607	0.6120
	SSIM	0.5838	0.6712	0.7694	0.8628	0.9335	0.9335	0.8124	0.7330	0.6750	0.6290
F-MNIST	MSE	0.5355	0.4298	0.3368	0.2657	0.2236	0.2236	0.2775	0.3255	0.3635	0.4026
	SSIM	0.4926	0.5989	0.6976	0.7753	0.8242	0.8242	0.7506	0.6835	0.6351	0.5914
CIFAR-10	MSE	0.3993	0.3439	0.2741	0.2061	0.1593	0.1593	0.2066	0.2420	0.2703	0.2920
	SSIM	0.4711	0.5429	0.6342	0.7289	0.8002	0.8002	0.7188	0.6619	0.6179	0.5883

TABLE I: Reconstruction quality vs sparsity and ambiguation levels.

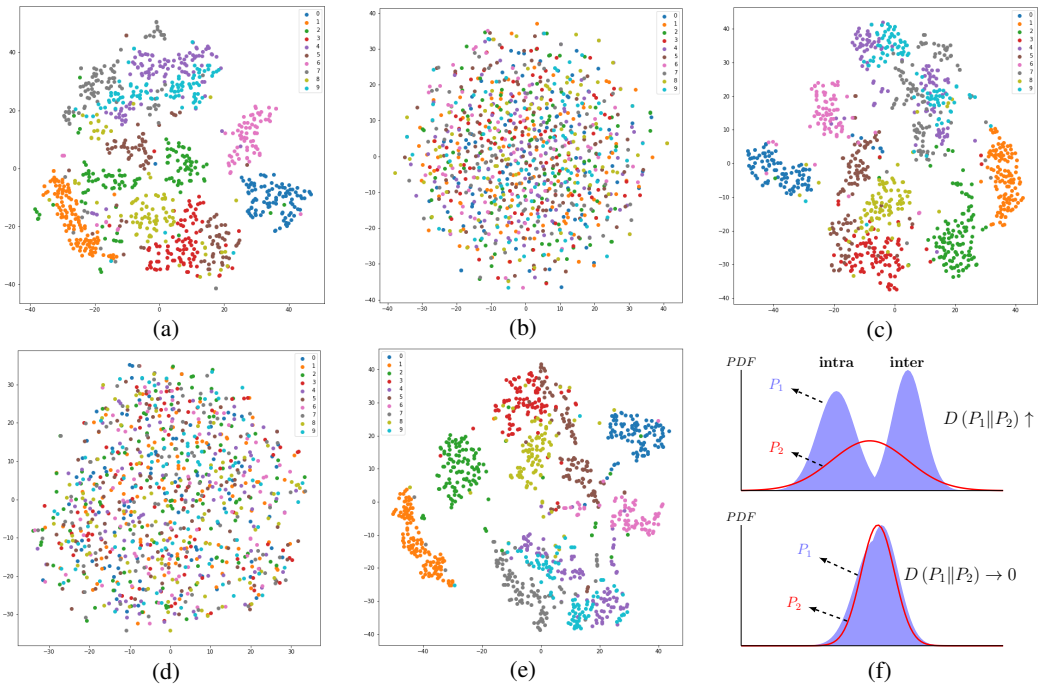

Fig. 4: t-SNE visualizations from MNIST dataset on: a) original space, b) transformed space with ambiguation, c) transformed space after purification, d) reconstructed space without knowledge of support, e) reconstructed space with knowledge of support; f) conceptual visualization of data clustering leakage.

terminology 'half ambiguation' is defined as $S_p = 0.5(L - S_x)$, and the terminology 'full ambiguation' is defined as $S_p = L - S_x$. Note that the information security guarantee addressed in [13], required keeping the projection parameters secretly.

Table I provides a quantitative comparison on reconstructed images using normalized MSE and SSIM (Structural Similarity Index) on MNIST [28], Fashion-MNIST [29], and CIFAR-10 [30] databases, where we applied the proposed method on latent representation of a designed convolutional autoencoder in [12], setting $L = 128$ and considering one code-map for MNIST and Fashion-MNIST databases and four code-maps for CIFAR-10 database. Finally, note that based on these results, our model follows the notion of (β, γ)-recoverable privacy mechanism, which we defined in Section II.

As a large-scale retrieval experiment, Figs. 3e-3f depict the recall measure for the CelebA database. The ground-truth was the pixel domain Euclidean distances and the latent code of the network in [12] is used to measure the approximate distances.

B. Clustering Leakage

Another potential threat is that the adversary may establish links between the closet disclosed representations. We now discuss database clustering leakage under our model. Note that the proposed mechanism can apply to privacy-preserving clustering applications where the goal is to perform clustering without disclosing the original data. The significant benefit of our method is that the authorized data users can purify the imposed ambiguation noise. However, the adversary will face a combinatorial problem to guesses the correct components.

Fig. 4 provides a qualitative visualization of clustering leakage on MNIST database [28], for which t-distributed stochastic neighbor embedding (t-SNE) [31] is used to project the underlying space to 2D. As illustrated, our model prevent database clustering leakage. Denoting by P_{intra} and P_{inter} as probability density functions of 'intra-cluster' and 'inter-cluster' of distances, respectively, Fig. 4f, provides a conceptual visualization of database clustering leakage, where $D(P_1\|P_2) = \mathbb{E}_{P_1}[\log \frac{P_1}{P_2}]$.

V. Conclusion

We present a computationally efficient, fairness-aware privacy-preserving nearby search scheme that can be utilized in cloud-based applications. The key insight behind our mechanism is that by approximating sparse representation of data points and adding random noise to their orthogonal complement, we can control privacy and utility trade-off in terms of dataset reconstruction and dataset clustering. The authorized data users can purify the ambiguated public representation thanks to the knowledge of correct support of the query.

References

[1] Jingdong Wang, Ting Zhang, Nicu Sebe, Heng Tao Shen, et al., "A survey on learning to hash," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 769–790, 2017.

[2] Piotr Indyk and Rajeev Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*. ACM, 1998, pp. 604–613.

[3] Aristides Gionis, Piotr Indyk, Rajeev Motwani, et al., "Similarity search in high dimensions via hashing," in *Vldb*, 1999, vol. 99, pp. 518–529.

[4] Jun Wang, Wei Liu, Sanjiv Kumar, and Shih-Fu Chang, "Learning to hash for indexing big data—a survey," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 34–57, 2015.

[5] Mayur Datar, Nicole Immorlica, Piotr Indyk, and Vahab S Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Annual Symposium on Computational Geometry*. ACM, 2004, pp. 253–262.

[6] Tobias Christiani, "Fast locality-sensitive hashing frameworks for approximate near neighbor search," in *Int. Conf. on Similarity Search and Applications*. Springer, 2019, pp. 3–17.

[7] Sariel Har-Peled and Sepideh Mahabadi, "Near neighbor: Who is the fairest of them all?," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 13176–13187.

[8] Ruslan Salakhutdinov and Geoffrey Hinton, "Semantic hashing," *International Journal of Approximate Reasoning*, vol. 50, no. 7, pp. 969–978, 2009.

[9] Yair Weiss, Antonio Torralba, and Rob Fergus, "Spectral hashing," in *Advances in neural information processing systems*, 2009, pp. 1753–1760.

[10] Sohrab Ferdowsi, Slava Voloshynovskiy, Dimche Kostadinov, and Taras Holotyak, "Sparse ternary codes for similarity search have higher coding gain than dense binary codes," in *IEEE Int. Symp. on Inf. Theory (ISIT)*, 2017.

[11] Behrooz Razeghi and Slava Voloshynovskiy, "Privacy-preserving outsourced media search using secure sparse ternary codes," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1–5.

[12] Sohrab Ferdowsi, Behrooz Razeghi, Taras Holotyak, Flavio P. Calmon, and Slava Voloshynovskiy, "Privacy-preserving image sharing via sparsifying layers on convolutional groups," in *IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[13] Petros Boufounos and Shantanu Rane, "Secure binary embeddings for privacy preserving nearest neighbors," in *IEEE Int. Work. on Inf. Forensics and Security (WIFS)*, 2011, pp. 1–6.

[14] Li Weng, Laurent Amsaleg, and Teddy Furon, "Privacy-preserving outsourced media search," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2738–2751, 2016.

[15] Krishnaram Kenthapadi, Aleksandra Korolova, Ilya Mironov, and Nina Mishra, "Privacy via the johnson-lindenstrauss transform," *arXiv preprint arXiv:1204.2606*, 2012.

[16] Shantanu Rane and Petros T Boufounos, "Privacy-preserving nearest neighbor methods: Comparing signals without revealing them," *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 18–28, 2013.

[17] Hervé Jégou, Matthijs Douze, and Cordelia Schmid, "On the burstiness of visual elements," in *IEEE Conf. on Comp. Vision and Pattern Recog. (CVPR)*, 2009, pp. 1169–1176.

[18] Florent Perronnin and Christopher Dance, "Fisher kernels on visual vocabularies for image categorization," in *IEEE Conf. on Comp. Vision and Pattern Recog. (CVPR)*, 2007, pp. 1–8.

[19] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez, "Aggregating local descriptors into a compact image representation," in *IEEE Conf. on Comp. Vision and Pattern Recog. (CVPR)*, 2010, pp. 3304–3311.

[20] Artem Babenko, Anton Slesarev, Alexandr Chigorin, and Victor Lempitsky, "Neural codes for image retrieval," in *Europ. Conf. on Comp. Vision*. Springer, 2014, pp. 584–599.

[21] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations (ICLR)*, 2014.

[22] Saiprasad Ravishankar and Yoram Bresler, "Learning sparsifying transforms," *IEEE Trans. on Signal Processing*, vol. 61, no. 5, pp. 1072–1086, 2013.

[23] Marzieh Gheisari, Teddy Furon, and Laurent Amsaleg, "Joint learning of assignment and representation for biometric group membership," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 2922–2926.

[24] Marzieh Gheisari, Teddy Furon, and Laurent Amsaleg, "Group membership verification with privacy: Sparse or dense?," in *2019 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2019, pp. 1–7.

[25] Marzieh Gheisari, Teddy Furon, Laurent Amsaleg, Behrooz Razeghi, and Slava Voloshynovskiy, "Aggregation and embedding for group membership verification," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 2592–2596.

[26] Dimche Kostadinov, Slava Voloshynovskiy, and Sohrab Ferdowsi, "Learning overcomplete and sparsifying transform with approximate and exact closed form solutions," in *European Workshop on Visual Information Processing*, 2018.

[27] Behrooz Razeghi, Slava Voloshynovskiy, Dimche Kostadinov, and Olga Taran, "Privacy preserving identification using sparse approximation with ambiguization," in *IEEE Int. Work. on Info. Forensics and Security (WIFS)*, 2017, pp. 1–6.

[28] Yann LeCun and Corinna Cortes, "MNIST handwritten digit database," 2010.

[29] Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[30] Alex Krizhevsky, Geoffrey Hinton, et al., "Learning multiple layers of features from tiny images," 2009.

[31] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.