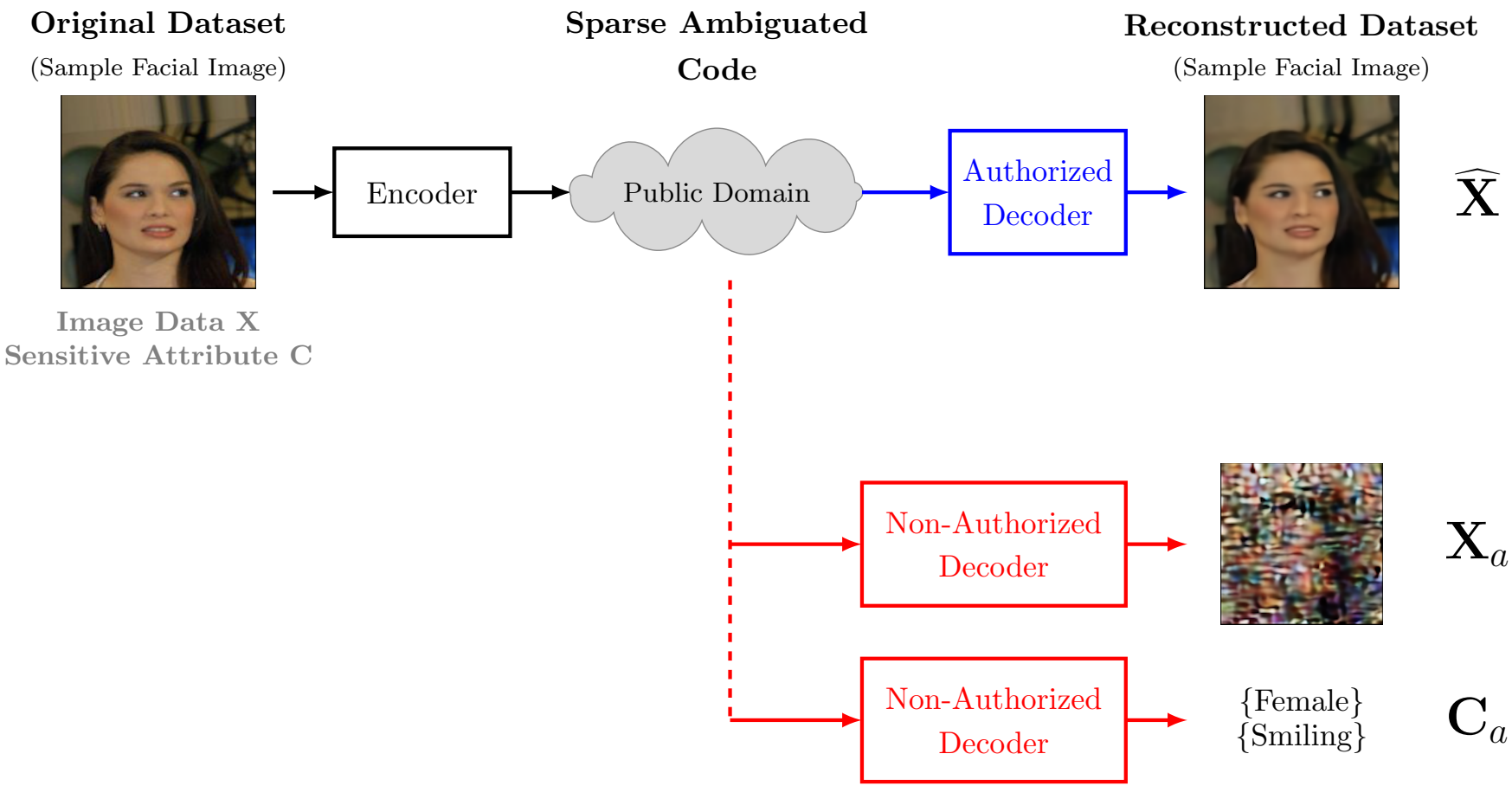
Compressed Data Sharing based on Information Bottleneck Model

Behrooz Razeghi, Shideh Rezaeifar, Sohrab Ferdowsi,
Taras Holotyak, Slava Voloshynovskiy

Department of Computer Science, University of Geneva

ICASSP'22 (virtual presentation)
May 2022

Motivation: Secure Data Sharing



General Setup

- Given the observed data $\mathbf{X} \sim p(\mathbf{X})$
- **Data Utility Objective**: Release a compact public representation $\mathbf{Z}_{\mathrm{p}}$ to provide some utility service (e.g., recognition, reconstruction, etc.) for some authorized parties.
- **Data Privacy Objective:** Restrict an adversary's eavesdropping on the released representation.
- **Our Objetive** Data-driven obfuscation mechanism smoothly trades-off the **informativeness** of the **bottleneck latent representation** for the **utility** service at hand against the **compressiveness** of the bottleneck variable from original (private) data, while at the same time minimizes the **privacy** leakage.

Problem Formulation

- Given the observed data $\mathbf{X}$
- **Objective:** Find the private bottleneck representation $\mathbf{Z}$ and its public counterpart $\mathbf{Z}_{\mathrm{p}}$ such that given another random variable $\mathbf{K}$, playing a role of a secret key available for the authorized parties only, we have

$$\Pr\{\mathbf{Z}_{\mathrm{p}} \mid \mathbf{K}\} \approx \Pr\{\mathbf{Z}\}$$

- The problem can be formulated by minimization of Lagrangian:

$$\mathcal{L}^{\mathrm{d}} = I(\mathbf{X}; \mathbf{Z}_{\mathrm{p}} \mid \mathbf{K}) - \beta I(\widehat{\mathbf{X}}; \mathbf{Z}_{\mathrm{p}} \mid \mathbf{K}),$$

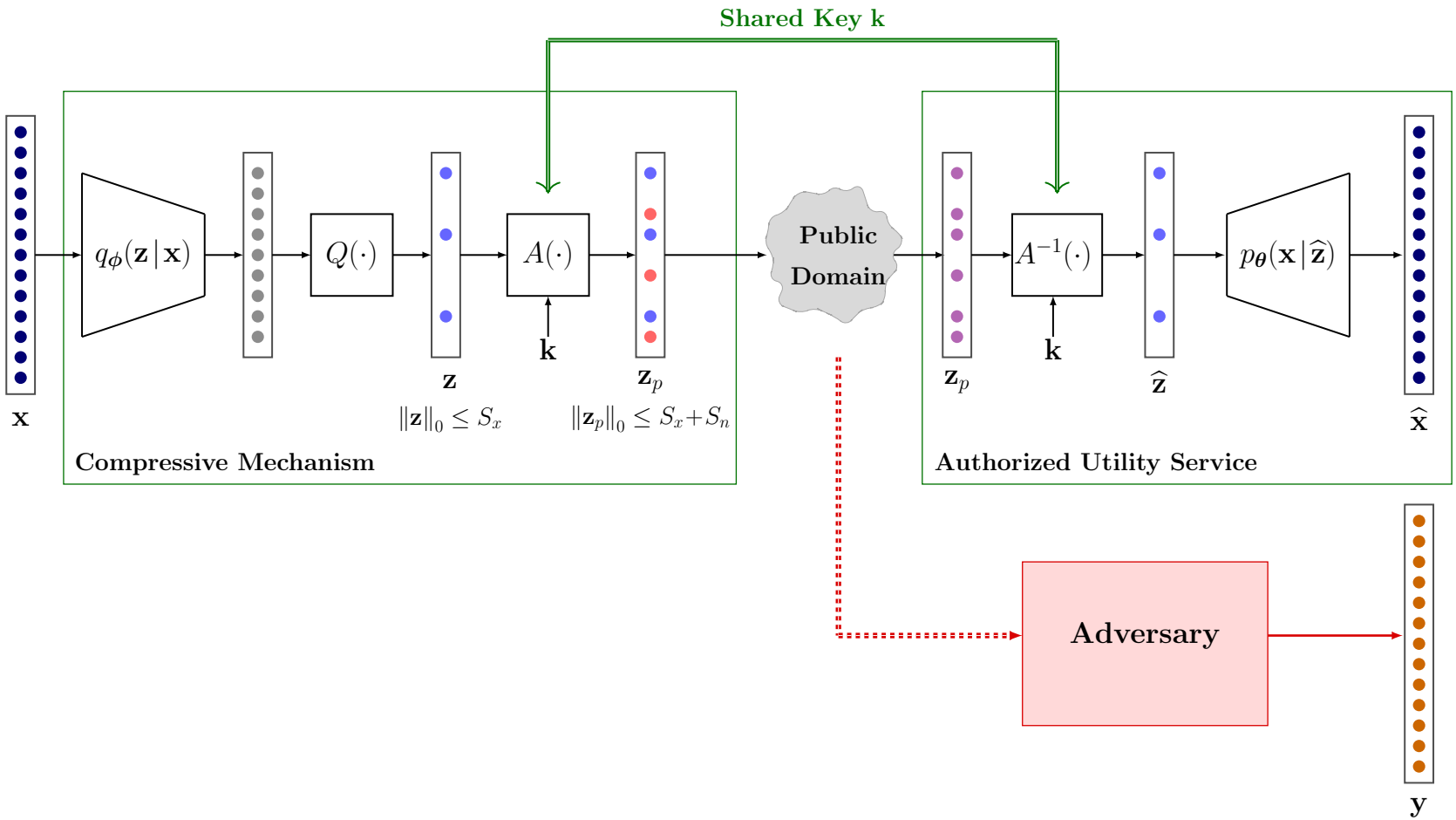
  on the side of defender.
- If the utility service is a reconstruction task, the *defender* minimization Lagrangian functional can be written as:

$$\mathcal{L}^{\mathrm{d}}\left(\boldsymbol{\theta}, \boldsymbol{\phi}\right) = I_{\boldsymbol{\phi}}\left(\mathbf{X}; \mathbf{Z}\right) - \beta I^{L}_{\boldsymbol{\theta}, \boldsymbol{\phi}}\left(\mathbf{Z}; \mathbf{X}\right),$$

  leading to the minimization problem:

$$\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\phi}}\right) = \underset{(\boldsymbol{\theta}, \boldsymbol{\phi})}{\arg\min}\, \mathcal{L}^{\mathrm{d}}\left(\boldsymbol{\theta}, \boldsymbol{\phi}\right).$$

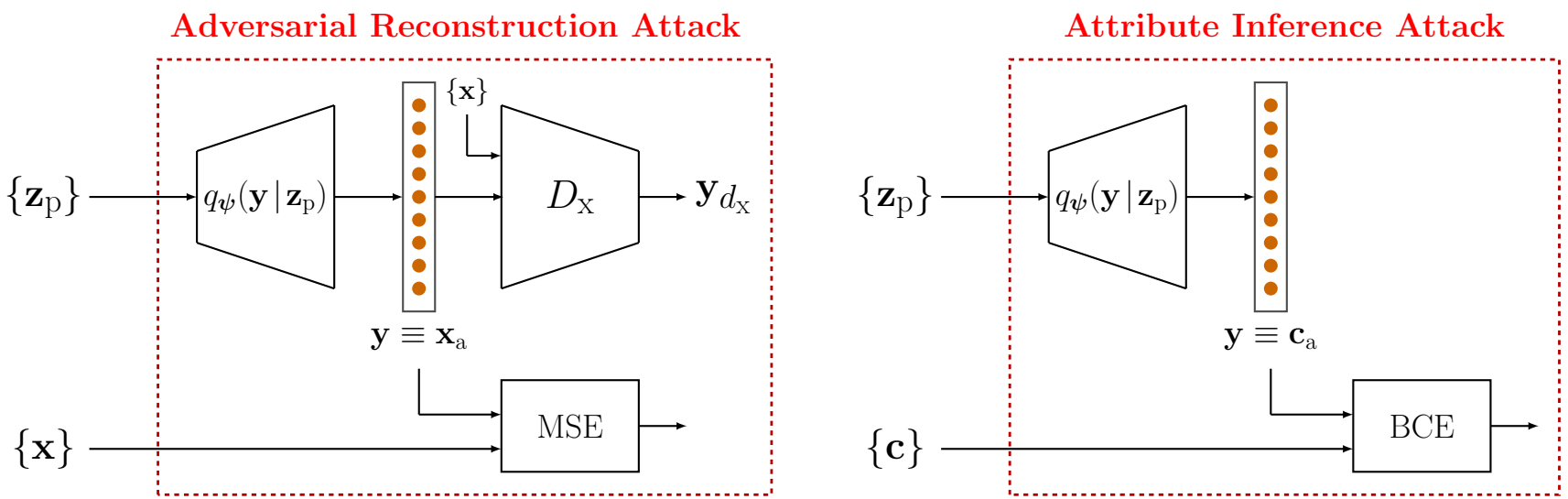
Operational Setup of the Proposed Mechanism



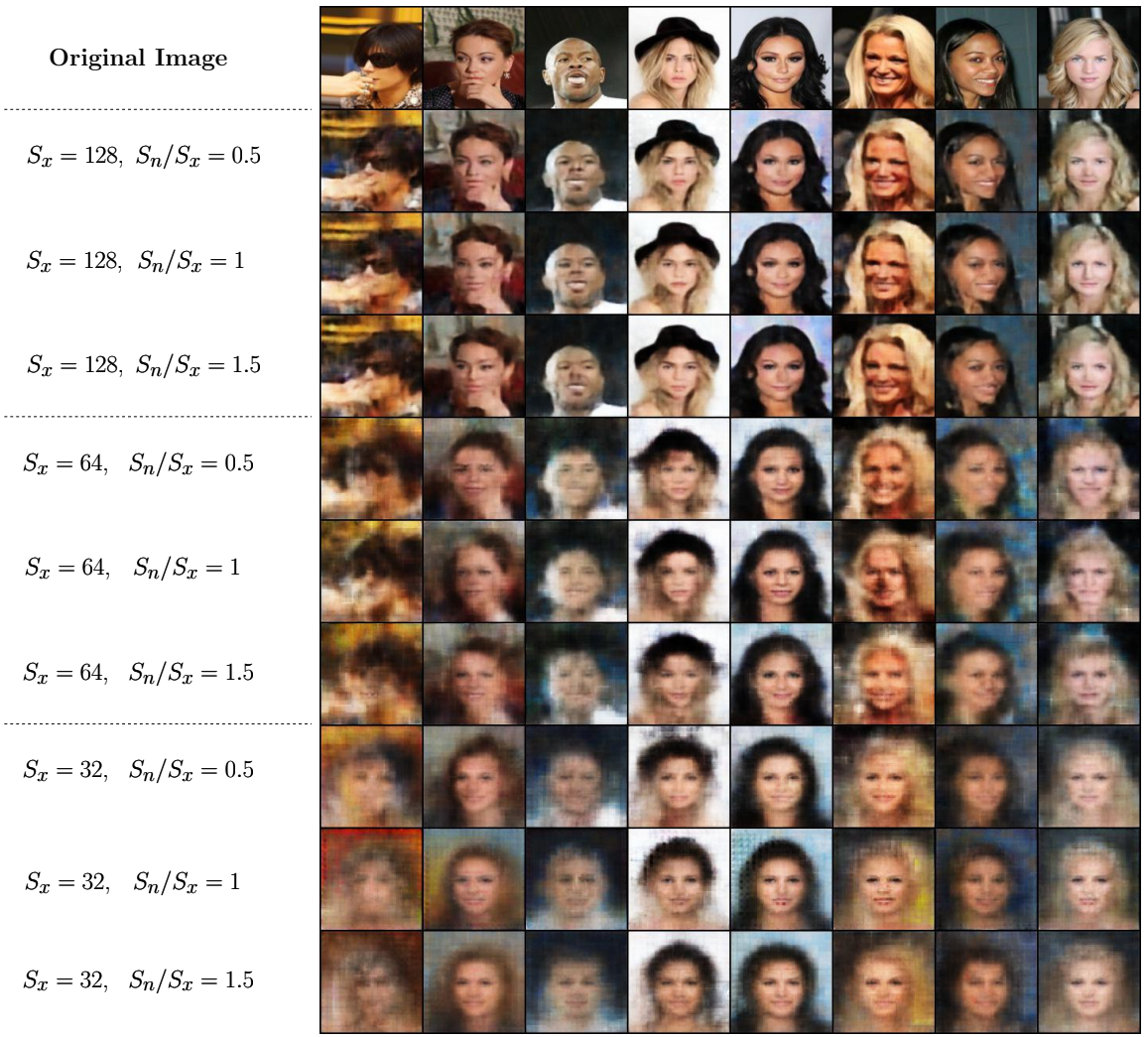
Threat Model

Assumptions:

- Adversary observes the compressed and protected public data $\mathbf{Z}_{\mathrm{p}}$
- Attacker knows the algorithm of encoding (compliant with Kerkhoff's principle)
- Secret key used for the ambiguation is unknown
- Attacker has access to a collection of training set



Visual Performance of Reconstruction Attack



Performance of Attribute Inference Attack

Classification Accuracy for CelebA dataset

Attribute	Sparsity Level	Ratio $\frac{S_n}{S_x}$			
	S_x	0	0.5	1	1.5
Gender	128	94.69	93.72	93.60	93.52
Gender	64	94.35	89.68	88.78	86.78
Gender	32	93.26	64.61	62.22	58.71
Age	128	85.028	84.46	83.20	82.98
Age	64	83.75	82.15	81.61	81.17
Age	32	83.08	78.44	77.48	77.42

Conclusions

- We demonstrate that the considered sparse coding with ambiguisation mechanism can ensure the secure data sharing in face of investigated adversarial reconstruction attack.
- However, while the reconstructed images are unrecognizable for the human observer under the proper ambiguization parameters, the trained adversarial classifier can still reliably infer the information about the sensitive attributes from the protected public representations.

$$H(\mathbf{X}) \gg H(\mathbf{C})$$

- That is why there is a need in the design of special encoding strategies ensuring both protection against unauthorized reconstruction attack and **known** targeted sensitivity attribute's attack. This is a topic of our ongoing work.