arXiv:2404.12170

Secure Semantic Communication for Image Transmission in the Presence of Eavesdroppers

Shunpu Tang, Chen Liu, Qianqian Yang, Shibo He, Dusit Niyato

https://arxiv.org/abs/2404.12170

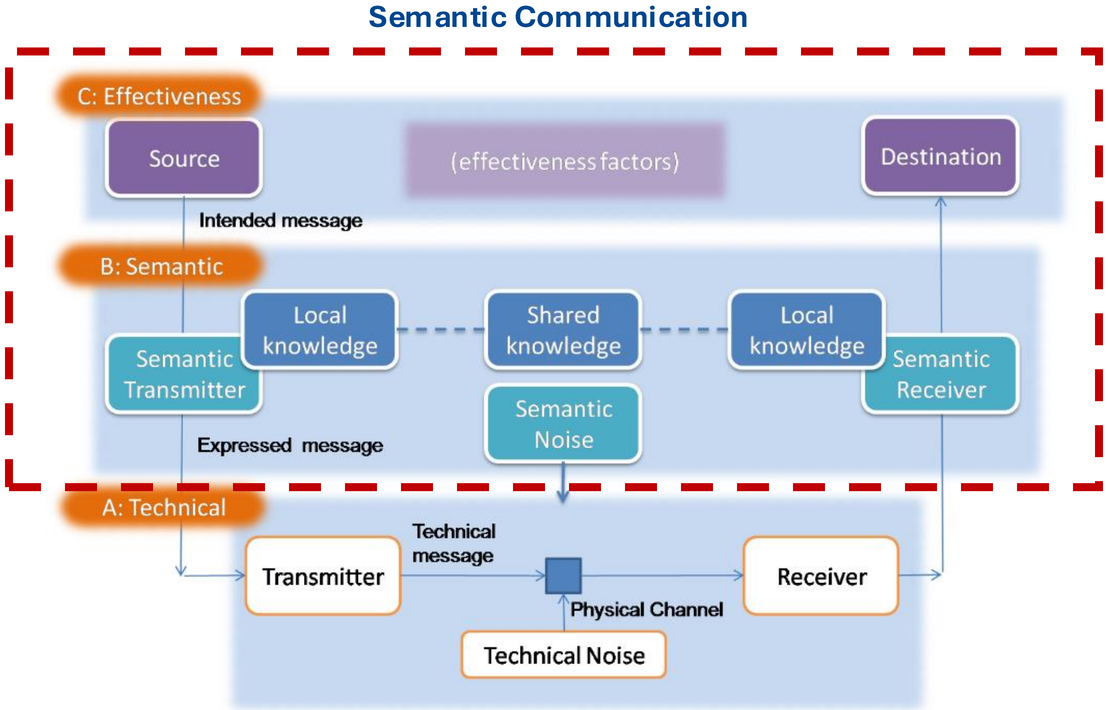
Cape Town • South Africa

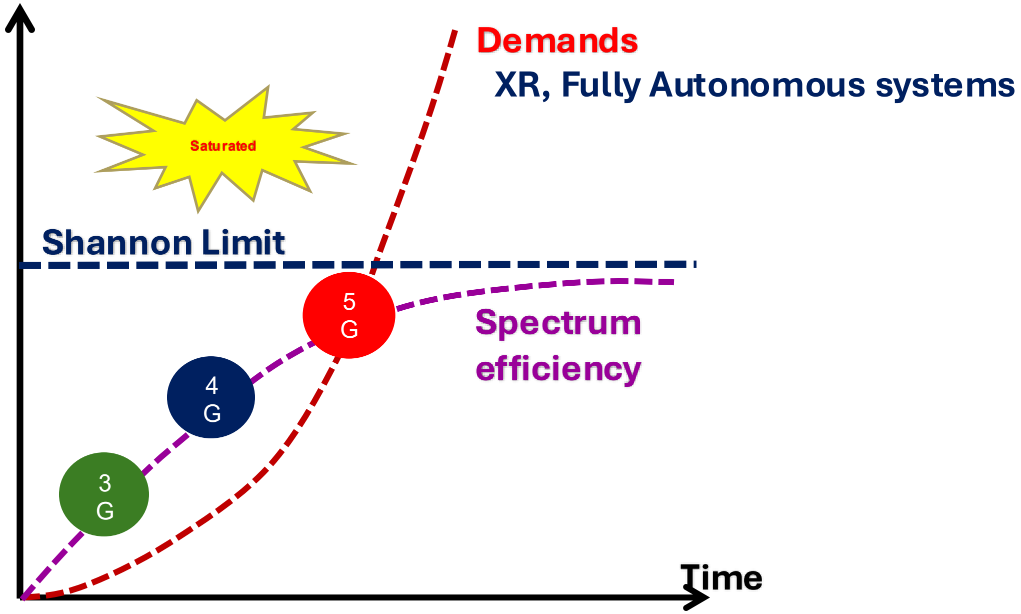
Outline

- Background
- Proposed System
- Simulation Results
- Conclusion

Backgrounds: SemCom

- **Semantic Communication (SemCom) is a promising technology for future network:**
  - **focuses on transmitting the meaning of the data rather than just bits.**
  - **aligns well with the increasing demands of machine-to-machine (M2M) communication**





[1] C.E. Shannon and W. Weaver, "The mathematical theory of communication," The University of Illinois Press, 1949.

Backgrounds: SemCom

Digital communication

10100100111
00010010010
01010101001
0010100111
…

Semantic encoding

Semantic information

落霞与孤鹜齐飞 秋水共长天一色 [2]

Solitary widgeon flies high in the sky, with rosy clouds floating aside; the river runs far near the horizon, reflecting the same color of sky

Les couchers de soleil volent avec les oiseaux solitaires, les eaux d'automne se confondent avec le ciel sans fin en une seule couleur

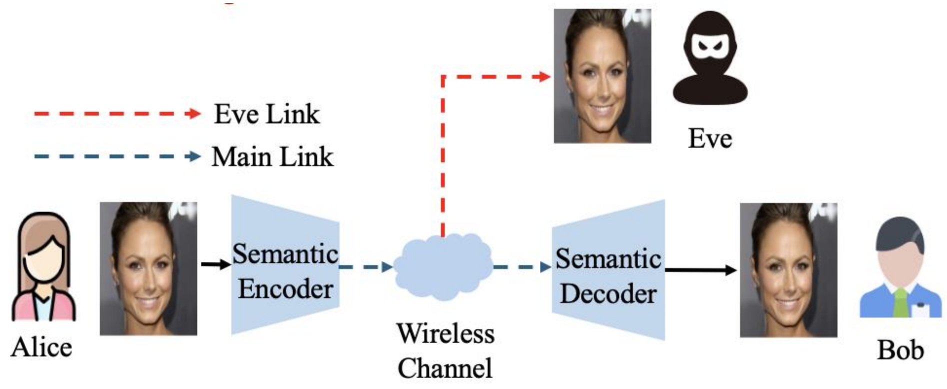
Las puestas de sol vuelan juntas con las aves solitarias, las aguas otoñales se funden con el cielo interminable en un solo color

Semantic decoding

[2] Bo Wang, "Preface to the Pavilion of King Teng Wang" in Ancient Chinese Literature, Tang Dynasty, A.D. 675

Background

■ SemCom is vulnerable to eavesdropping



The open nature of the wireless channel enables Eve to capture the transmitted signals!

□ High Interpretability of Semantic Data:

- SemCom transmits task-relevant semantic information, making the data more interpretable.

□ Challenges in Applying Classical Security Approaches:

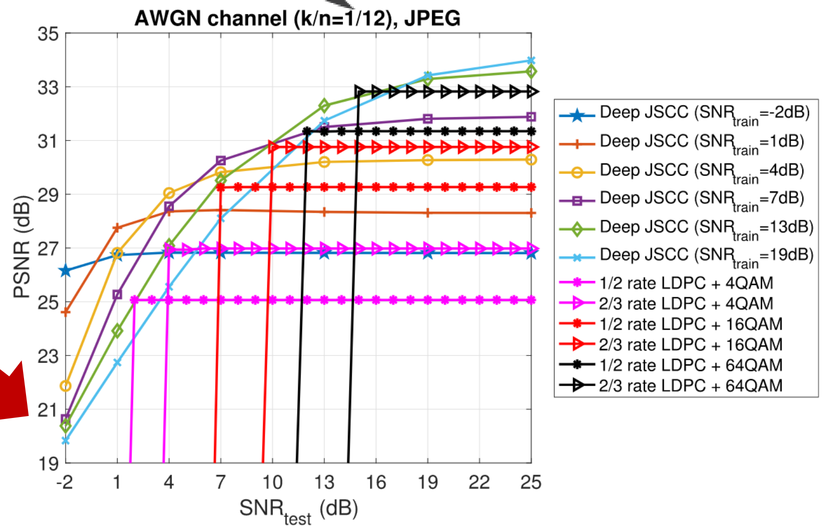
- SemCom typically bypass quantization (Joint source-channel coding).
- Even if the eavesdropper's channel is very worse, it may decode successfully (powerful semantic decoder)

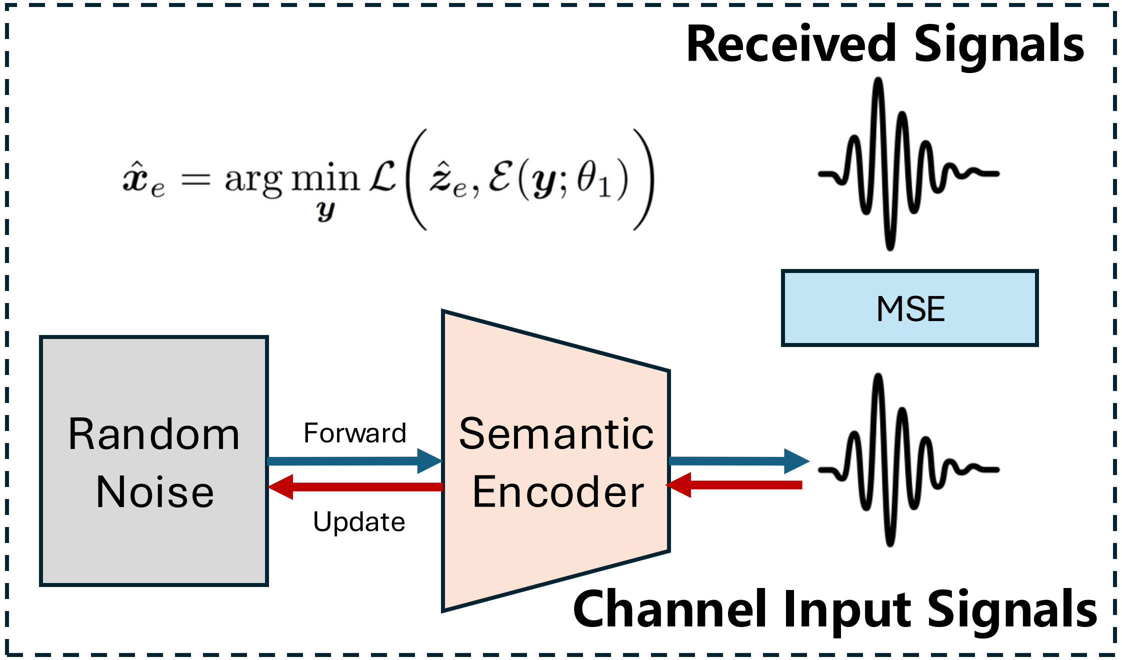
× Encryption: AES, SNOW 3G, ZUC
× Physical layer security: Modulation and Coding, Resource Allocation, Beamforming

Background

■ Eavesdropping Methods in SemCom

□ **Naive Decoding**: **If Eve gains access to the semantic decoder, it may decode sensitive semantic information, even when its channel conditions are significantly worse than that of Bob.**



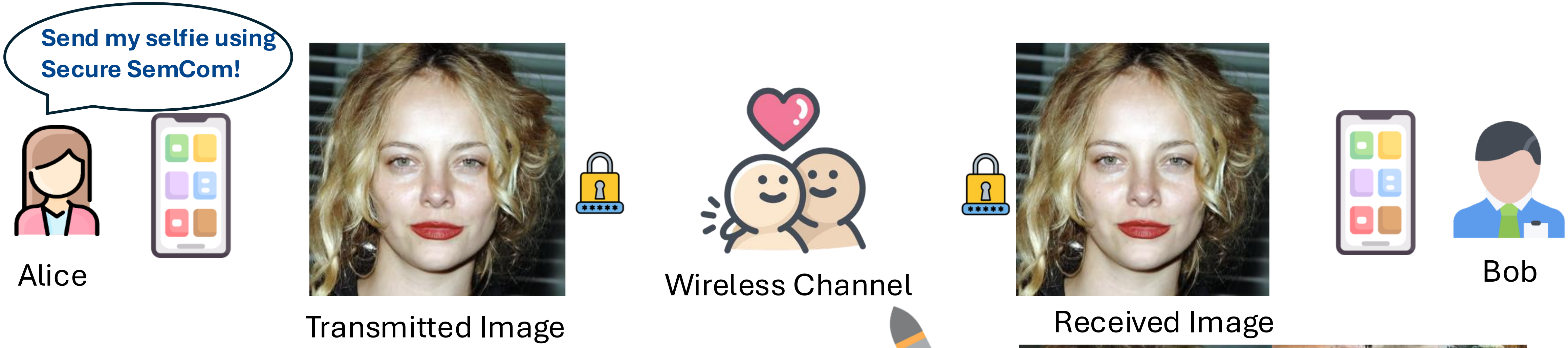


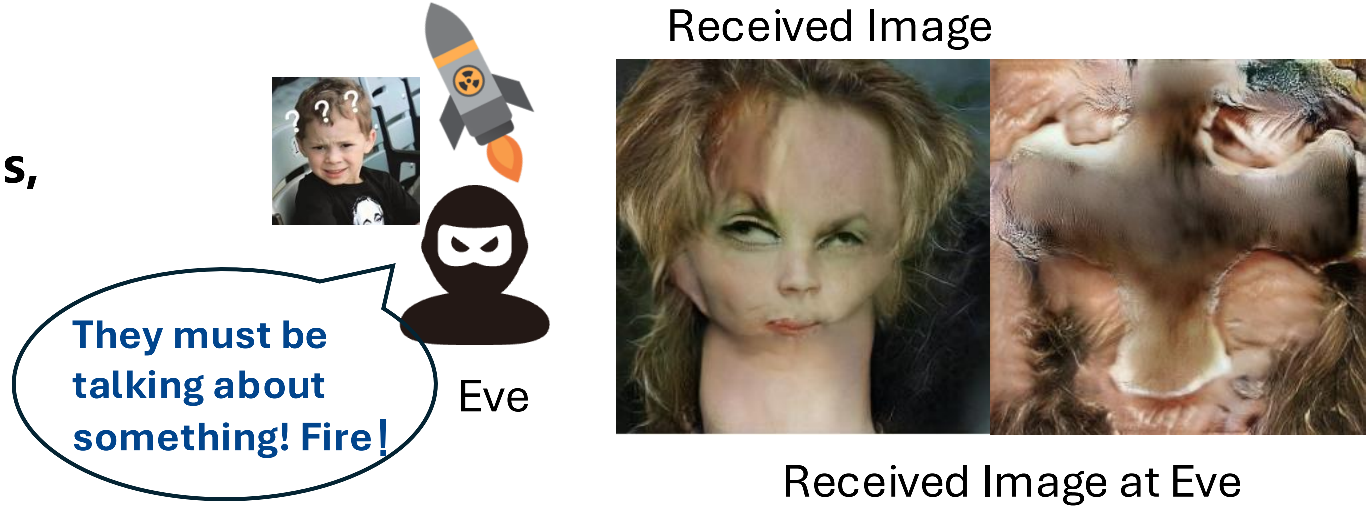
□ **Model Inversion Attacks (MIA)**: **Eve can attempt to reverse-engineer the semantic encoder to extract private information, when it can only access to the semantic encoder.**

[1] Y. Chen, et. al. “The model inversion eavesdropping attack in semantic communication systems” IEEE GlobeCom 2023

Background

■ Existing Works in Secure SemCom

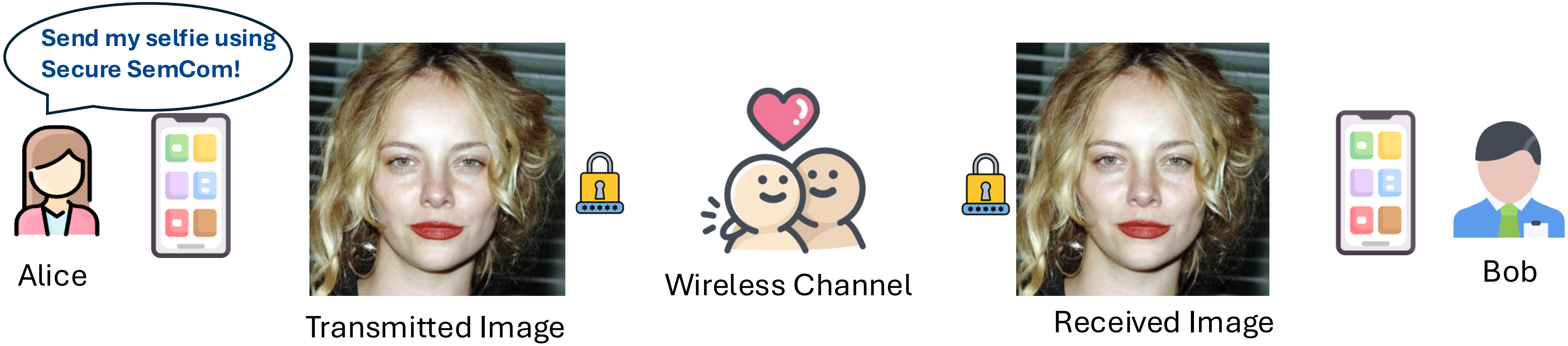




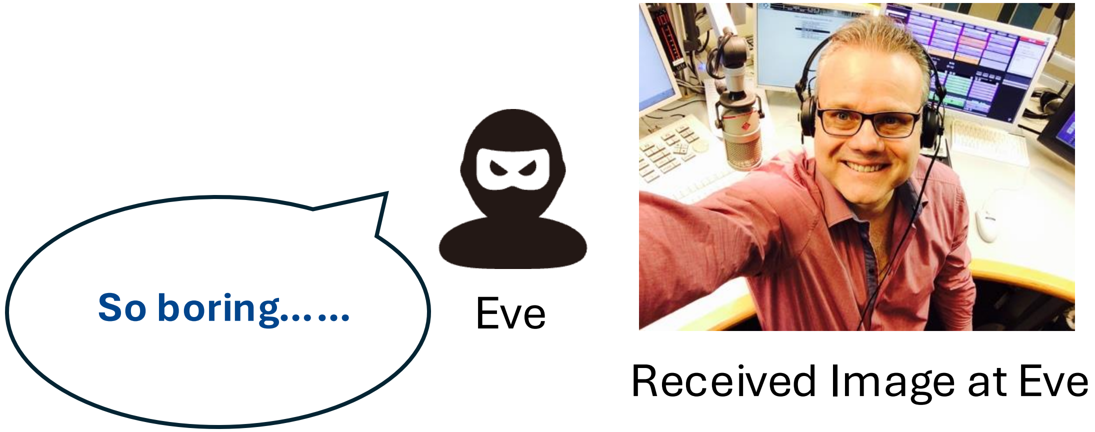
- ◻ Existing methods like loss functions, encryption schemes, effectively reduce the risk of leakage.
- ◻ However, it may Raise Suspicions from Eve, prompt Eve to exert jamming attacks to corrupt communication.

Background

■ Our Motivation

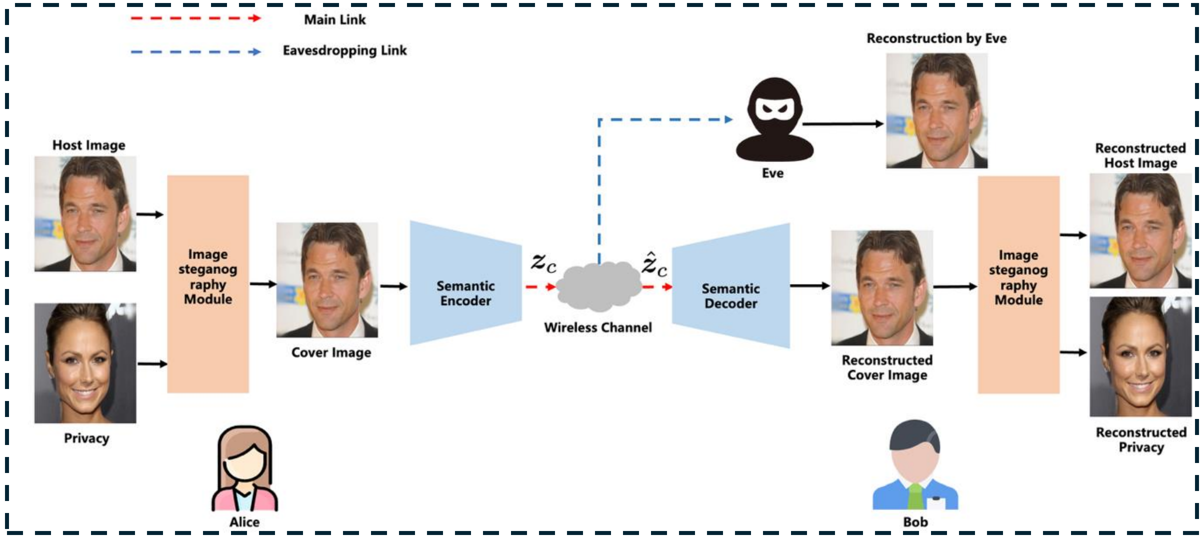


□ **Inspired by the concept of covert communication, we aim to mislead them into believing the data is irrelevant or meaningless.**



Proposed System

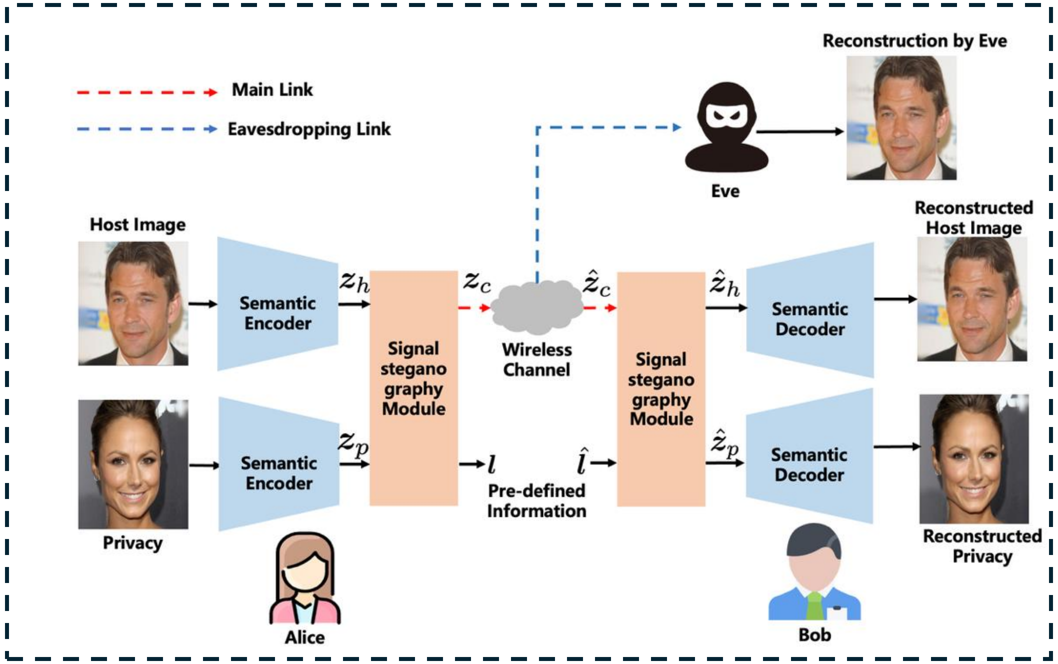
☹ Image Steganography



- ☐ A straightforward way is to directly exert steganography at the source images:
  - ✗ Pose challenges for the well-trained SemCom system
  - ✗ Deteriorate the reconstruction performance of the private image

☺ Signal Steganography

- ☑ We apply steganography on the channel input signal before transmission.
- ☑ Our goal is to mislead eavesdroppers by leveraging signal steganography to conceal the private image within any non-sensitive image.



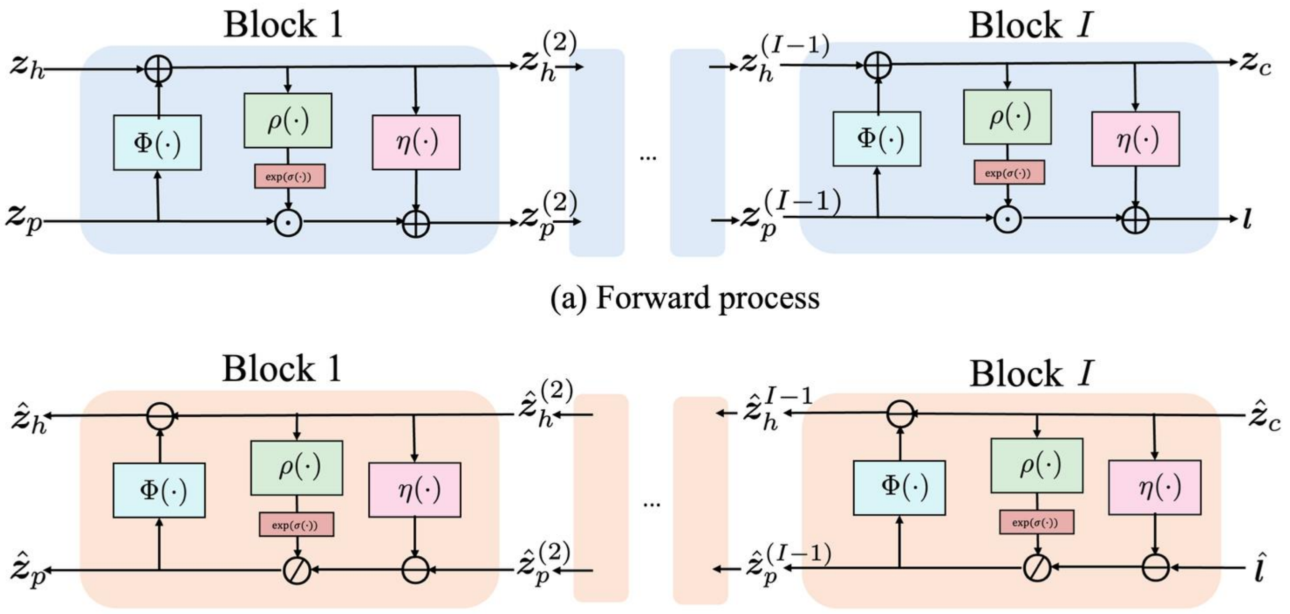
Proposed System

■ INN-based Signal Steganography Module

- ☐ **INN enables reversible operations, ensuring precise reconstruction of the original signals from the INN output signals.**
- ☐ **We propose a INN-based steganography module, which consists of invertible blocks with additive affine transformations.**
- ☐ **Only z_c is transmitted over the wireless channel, and $\hat{l}$ is a predefined constant value or sampled from a given distribution.**

$$\boldsymbol{z}_h^{(i+1)} = \boldsymbol{z}_h^{(i)} + \Phi(\boldsymbol{z}_p^{(i)}),$$

$$\boldsymbol{z}_p^{(i+1)} = \boldsymbol{z}_p^{(i)} \odot \exp\left(\rho(\boldsymbol{z}_h^{(i+1)})\right) + \eta(\boldsymbol{z}_h^{(i+1)})$$



(a) Forward process

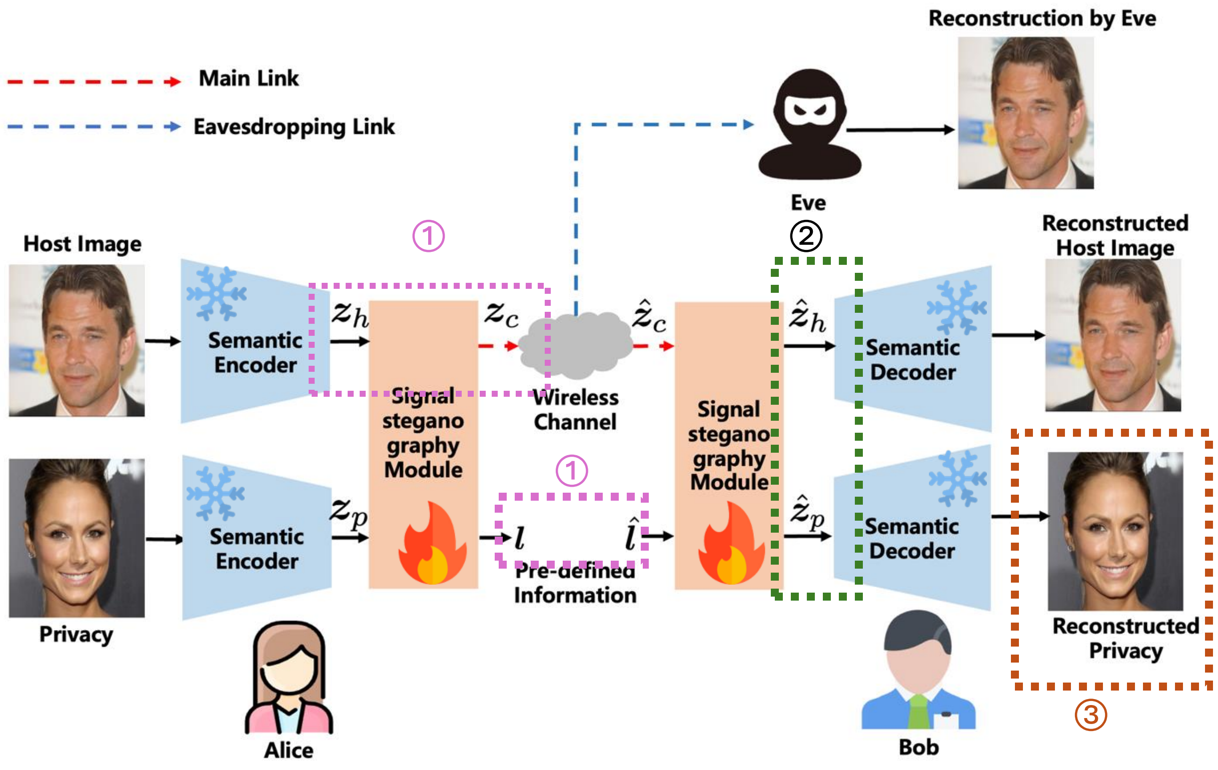
(b) Backward process

$$\hat{\boldsymbol{z}}_p^{(i)} = \left(\hat{\boldsymbol{z}}_p^{(i+1)} - \eta(\hat{\boldsymbol{z}}_h^{(i+1)})\right) \odot \exp\left(-\rho(\hat{\boldsymbol{z}}_h^{(i+1)})\right)$$

$$\hat{\boldsymbol{z}}_h^{(i)} = \hat{\boldsymbol{z}}_h^{(i+1)} - \Phi(\hat{\boldsymbol{z}}_p^{(i)}).$$

Proposed System

- Training procedure and Loss function



Total loss function

$$\mathcal{L}_{total} = \mathcal{L}_{\text{forward}} + \mathcal{L}_{\text{backward}} + \mathcal{L}_{\text{privacy}}$$

① Signal Steganography

$$\mathcal{L}_{\text{forward}} = \lambda_1 ||\boldsymbol{z}_c - \boldsymbol{z}_h||_2^2 + \lambda_2 ||\boldsymbol{l} - \hat{\boldsymbol{l}}||_2^2$$

② Signal Reconstruction

$$\mathcal{L}_{\text{backward}} = \lambda_3 ||\boldsymbol{z}_p - \hat{\boldsymbol{z}}_p||_2^2 + \lambda_4 ||\boldsymbol{z}_h - \hat{\boldsymbol{z}}_h||_2^2$$

③ Private image Reconstruction

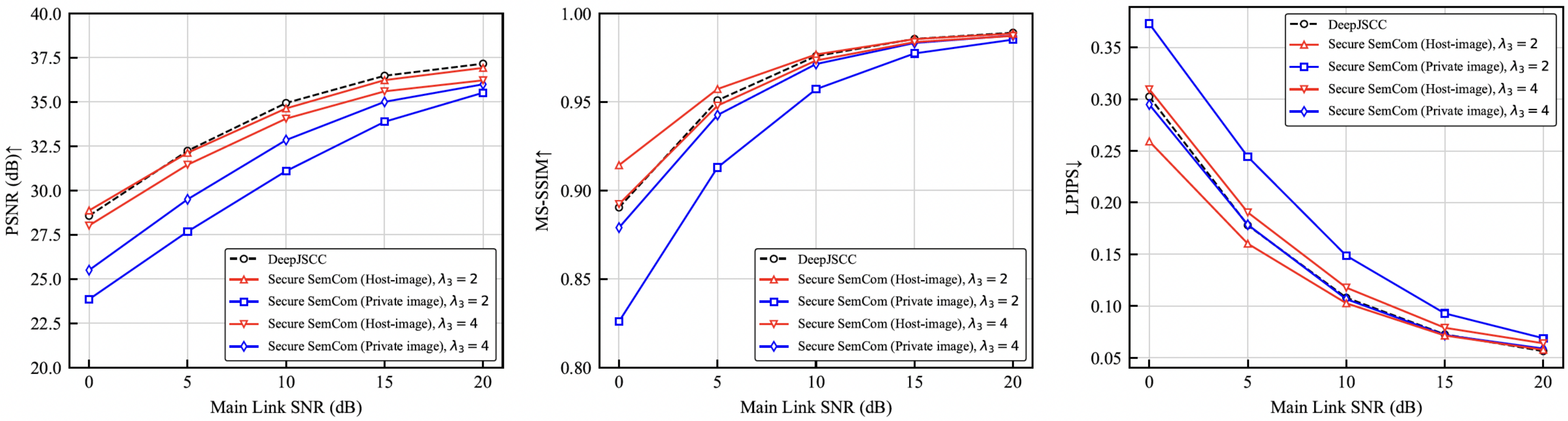
$$\mathcal{L}_{\text{privacy}} = \lambda_5 ||\boldsymbol{x}_p - \hat{\boldsymbol{x}}_p||_2^2$$

Simulation

■ Implementation Details

- ◻ Dataset:
  - CelebA-Masked-HQ
  - 2,500 random pairs of host and private images
- ◻ Semantic Encoder and Decoder:
  - Pre-trained DeepJSCC with a bandwidth compression ratio (BCR) of 1/12.
- ◻ Invertible Neural Network:
  - 8 invertible blocks.
- ◻ Baseline:
  - Transmits private images without secure mechanisms with DeepJSCC.
- ◻ Evaluation Metrics:
  - PSNR, MS-SSIM, LPIPS for Bob and Eve, respectively

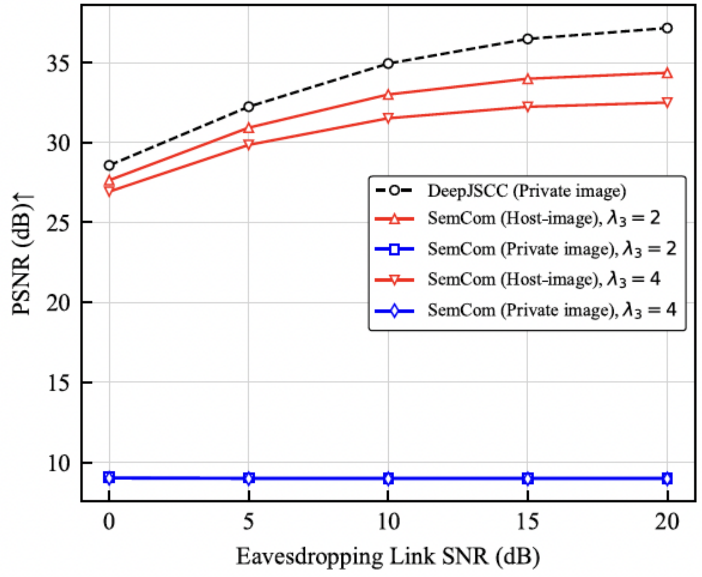
■ Reconstruction performance at Bob



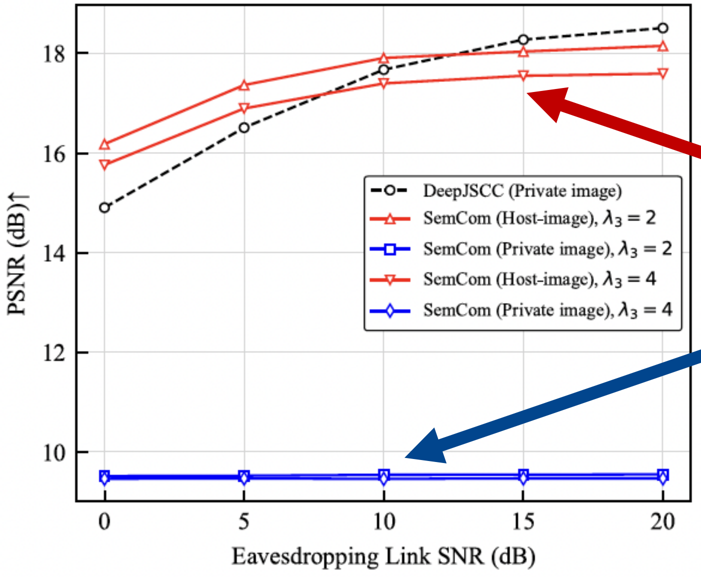
(a) (b) (c)

- ☐ **The proposed approach maintains comparable reconstruction quality of the private image compared to the scenario without any secure mechanisms**

■ Reconstruction performance at Eve



(a) Naive Decoding



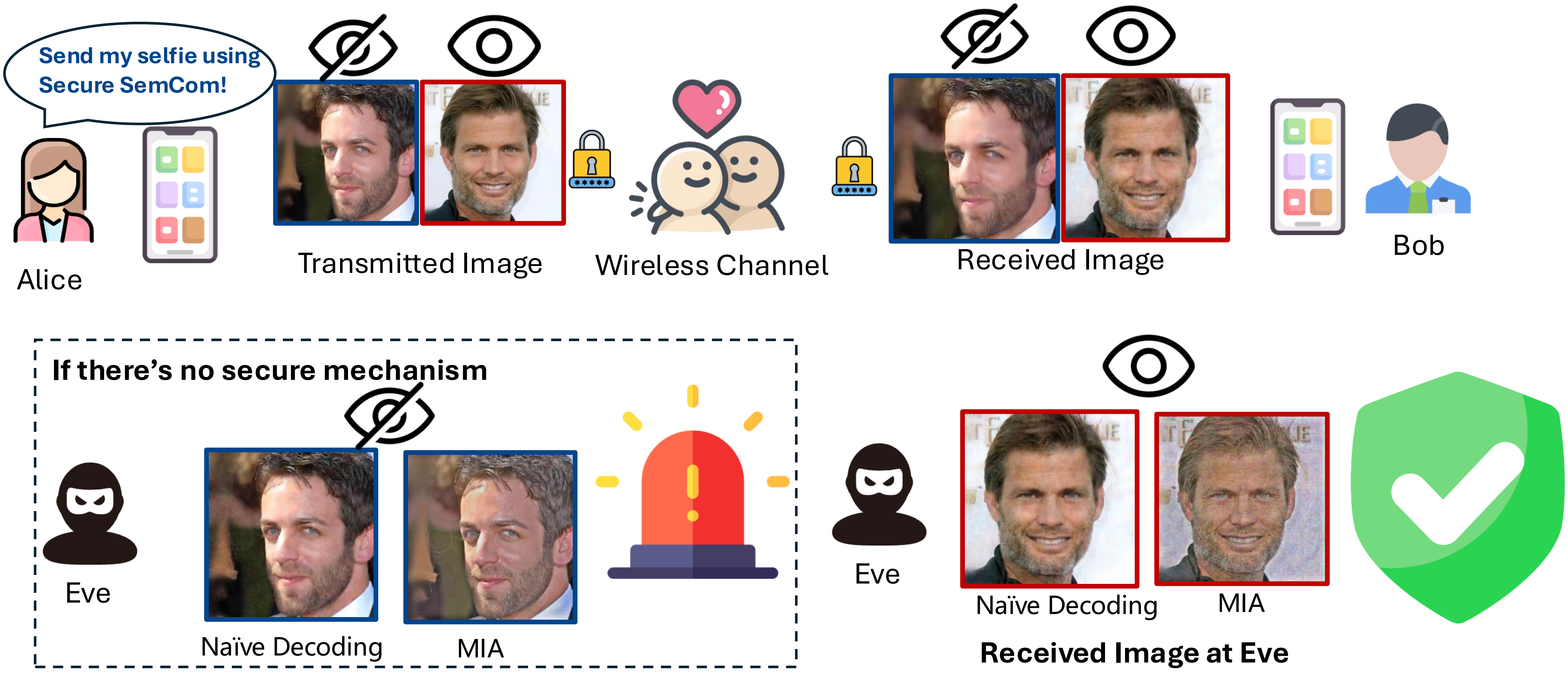
(b) MIA

	Naive Decoding		MIA	
Method	MS-SSIM	LPIPS	MS-SSIM	LPIPS
DeepJSCC (Private image)	0.951	0.178	0.604	0.627
Secure SemCom (Private image), λ_3=2	0.275	0.639	0.166	0.737
Secure SemCom (Private image), λ_3=4	0.270	0.639	0.160	0.744
Secure SemCom (host image), λ_3=2	0.943	0.195	0.686	0.584
Secure SemCom (host image), λ_3=4	0.936	0.212	0.663	0.602

- **Images reconstructed by Eve exhibit significant similarity to the original host image,**
- **and they are very different to the private images.**

- **In contrast, Eve can easily decode the private image by eavesdropping on the link with conventional DeepJSCC.**

■ Visual Comparison



Conclusion

IEEE GLOBECOM®

- ✓ Developed a novel invertible neural network (INN)-based steganography module to embed private signals within host signals, concealing sensitive information during transmission.
- ✓ Ensured that the legitimate receiver (Bob) achieves comparable image quality to systems without security mechanisms.
- ✓ Demonstrated that the eavesdropper can only reconstruct host images, effectively protecting sensitive content.

Thanks!

Email: tangshunpu@zju.edu.cn

arXiv:2404.12170