STAR-RIS-Assisted Privacy Protection in Semantic Communication System

Yiru Wang, *Graduate Student Member, IEEE*, Wanting Yang, Pengxin Guan, Yuping Zhao, and Zehui Xiong, *Senior Member, IEEE*

***Abstract*—Semantic communication (SemCom) has emerged as a promising architecture in the realm of intelligent communication paradigms. SemCom involves extracting and compressing the core information at the transmitter while enabling the receiver to interpret it based on established knowledge bases (KBs). This approach enhances communication efficiency greatly. However, the open nature of wireless transmission and the presence of homogeneous KBs among subscribers of identical data type pose a risk of privacy leakage in SemCom. To address this challenge, we propose to leverage the simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) to achieve privacy protection in a SemCom system. In this system, the STAR-RIS is utilized to enhance the signal transmission of the SemCom between a base station and a destination user, as well as to convert the signal to interference specifically for the eavesdropper (Eve). Simulation results demonstrate that our generated task-level disturbance outperforms other benchmarks in protecting SemCom privacy, as evidenced by the significantly lower task success rate achieved by Eve.**

***Index Terms*—Semantic communication, task-oriented communication, privacy protection, STAR-RIS.**

I. Introduction

Recently, the industry and academia have witnessed the advances in artificial intelligence (AI), driving the shift from conventional communication to semantic communication (SemCom). SemCom, through the use of knowledge bases (KBs), enables the transmission of only vital semantic information relevant to the communication task [1]. The KBs contain relevant semantic knowledge of data and tasks, forming the foundation for establishing SemCom encoders and decoders [2]. However, in practice, users with the same data often share identical KBs, posing privacy risks in SemCom as unintended recipients might decode messages meant for others, especially in open wireless contexts.

In conventional communication, there have been some privacy protection technologies [3], [4], aiming to reduce the signal-to-noise ratio (SNR) at the eavesdropper (i.e. Eve). However, it has been demonstrated that SemCom can still achieve satisfactory performance even under unfavorable channel conditions [5], [6]. As a result, reducing the SNR at Eve is less effective in SemCom compared to conventional communication. In AI-based communication, evasion attacks can be launched to fool a user's model into making wrong decisions by manipulating its received data [7], which have been applied to wireless communication in terms of fooling the classifiers used for modulation recognition, channel estimation and many other areas [8], [9], [10]. However, evasion attacks cannot be directly generated to interfere with Eve and achieve privacy-preserving goal, considering that the destination user (i.e. Bob) may also suffer from those attacks due to the open characteristic of wireless channels. To regulate the efficiency-privacy trade-off between Bob and Eve, the authors in [11] trained a joint-source-and-channel (JSC) autoencoder to prevent Eve from cracking the semantic information. However, when the channel condition is comparably worse at Bob, the privacy protection performance will degrade.

To fill this gap, the simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) can be utilized to further enhance the privacy of the SemCom. STAR-RIS can divide the incident wireless signal into transmitted and reflected signal passing into both sides of the space surrounding the surface, thus facilitating a full-space manipulation of signal propagation [12]. There are three practical protocols for operating STAR-RIS in wireless communication systems, namely energy splitting (ES), mode switching (MS), and time switching (TS). In this work, the MS mode is selected since it is easier to implement and more spectrally efficient compared to ES and TS, respectively [13]. In MS protocol, partial elements are configured to fully-transmitting mode, while others are designed for fully-reflecting mode. We consider a scenario where Bob and Eve locate in the transmission and reflection region of the STAR-RIS, respectively. By carefully designing the transmission-coefficient vector (TCV) of the STAR-RIS, the signal can be effectively transmitted by the STAR-RIS to Bob. At the same time, the same incident signal can be adjusted by the STAR-RIS's reflection-coefficient vector (RCV) and reflected to form interference at Eve to degrade its task performance. By leveraging the STAR-RIS, the desired semantic signal and interference can be completely separated.

In this paper, we propose to use the STAR-RIS to realize privacy protection in SemCom-enabled services like online virtual reality meetings. We consider a scenario where a high-tech company, referred to Bob, operates in a room equipped with a STAR-RIS. The transmitted data, such as vehicle driving images and video views, is used for tasks like classification and recommendation. We aim to protect the SemCom system against a potential Eve, like competitive companies or other users in the network, by employing STAR-RIS for privacy protection. The main contributions are summarized as follows:

- We aim to achieve two objectives in this STAR-RIS-assisted SemCom system. One is to enhance the signal transmission of SemCom between the base station (BS) and Bob by jointly optimizing the TCV of STAR-RIS and the beamforming vector of the BS, while the other is to suppress the privacy leakage to Eve by designing STAR-RIS's RCV.
- We propose two methods for designing the RCV of STAR-RIS, by which the STAR-RIS's received signal is converted and reflected to form task-level and SNR-level interference at Eve. By

Manuscript received 25 June 2023; revised 31 January 2024; accepted 22 March 2024. Date of publication 1 April 2024; date of current version 19 September 2024. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1801403 and Grant 2018YFB1801402, in part by the National Natural Science Foundation of China under Grant 62271228, in part by the Jilin Scientific and Technological Development Program under Grant 20220101103JC, in part by the National Research Foundation, Singapore and Infocomm Media Development Authority through its Future Communications Research & Development Programme, in part by the SUTD under Grant SRG-ISTD-2021-165, in part by SUTD-ZJU IDEA under Grant SUTD-ZJU (VP) 202102, in part by the Ministry of Education, Singapore, through its SMU-SUTD Joint under Grant 22-SIS-SMU-048, and in part by SUTD Kickstarter Initiative under Grant SKI 20210204. The review of this article was coordinated by Prof. Hongzi Zhu. *(Corresponding author: Yuping Zhao.)*

Yiru Wang, Pengxin Guan, and Yuping Zhao are with the School of Electronics, Peking University, Beijing 100871, China (e-mail: yiruwang@stu.pku.edu.cn; guanpengxin@pku.edu.cn; yuping.zhao@pku.edu.cn).

Wanting Yang and Zehui Xiong are with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372 (e-mail: wanting_yang@sutd.edu.sg; zehui_xiong@sutd.edu.sg).

Digital Object Identifier 10.1109/TVT.2024.3383824

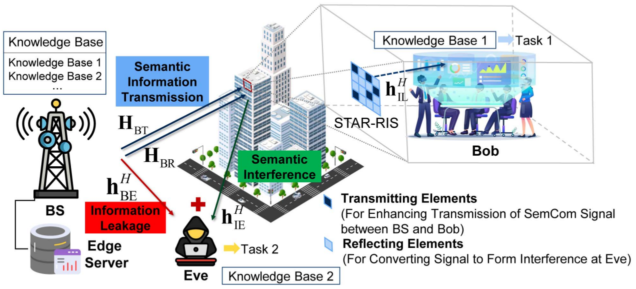

Knowledge Base
Knowledge Base 1
Knowledge Base 2
...
Semantic Information Transmission
$\mathbf{H}_{\mathrm{BT}}$
$\mathbf{H}_{\mathrm{BR}}$
$\mathbf{h}_{\mathrm{BE}}^{H}$
BS
Information Leakage
Edge Server
Semantic Interference
$\mathbf{h}_{\mathrm{IE}}^{H}$
Eve
Task 2
Knowledge Base 2
Knowledge Base 1
Task 1
$\mathbf{h}_{\mathrm{IL}}^{H}$
STAR-RIS
Bob
Transmitting Elements
(For Enhancing Transmission of SemCom Signal between BS and Bob)
Reflecting Elements
(For Converting Signal to Form Interference at Eve)

Fig. 1. STAR-RIS-assisted privacy protection in SemCom system.

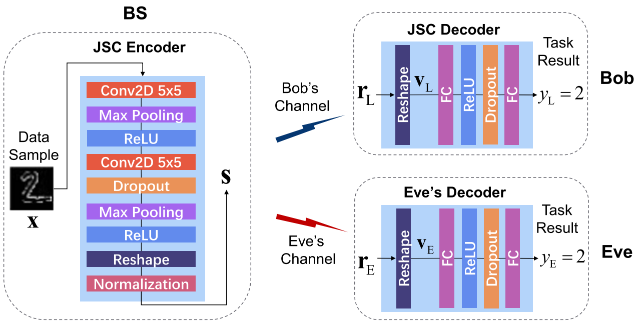


Fig. 2. Illustration of SemCom process and a possible privacy leakage.

leveraging STAR-RIS, the interference for Eve and the desired signal for Bob is separated, causing no disturbance to the SemCom between the BS and Bob.

- We evaluate the proposed methods through simulations. The results indicate that Eve exhibits a lower task success rate under task-level interference compared to SNR-level disturbance. Besides, our proposed schemes exhibit superiority over the existing methods, particularly when the channel condition of Eve is better than that of Bob. Furthermore, our approach for disturbing Eve is training-free, leading to reduced complexity and time delay.

Notations: x, $\mathbf{x}$, $\mathbf{x}^H$ and $\mathbf{X}$ and denote a scalar, a column vector, the conjugate transpose of the vector $\mathbf{x}$, and a matrix, respectively. $|x|$ and $\|\mathbf{x}\|_2$ denote the absolution value of a scalar x and Euclidean norm of a column vector $\mathbf{x}$. $\mathrm{Diag}(\mathbf{x})$ is a diagonal matrix with the entries $\mathbf{x}$ on its main diagonal.

The rest of this paper is organized as follows. Our system model is described in Section II. The proposed SemCom empowering and privacy protection mechanisms are introduced in Section III. After providing our simulation results in Section IV, we finally conclude in Section V.

II. System Model

The considered eavesdropping scenario is shown in Fig. 1. A STAR-RIS is deployed at the windows of the high-rise building to enhance the privacy of SemCom, which can be configured by the BS and Bob. Bob is in one room of the building, who locates in the transmission-region of the STAR-RIS. Eve locates outside the building and is much closer to the BS, who belongs to the reflection-region of the STAR-RIS.

A. Semantic Encoding and Decoding Models

In this paper, a task-oriented SemCom scenario is considered, where the encoder at the BS and the decoder at Bob are jointly trained in advance to achieve high classification performance. Eve independently trains its own decoder to align with the fixed encoder at the BS, aiming to maximize its task success rate. Hence, while Bob and Eve share the same encoder located at the BS, they each utilize their uniquely trained decoders. All communication participants are assumed to already exist, and hence the BS has stored all users' KBs. Privacy leakage occurs when the BS transmits a message intended for Bob but Eve eavesdrops it, as shown in Fig. 2.

We denote the image requested by Bob as $\mathbf{x} \in \mathbb{R}^{c \times h \times w}$. Different from the conventional separate source coding (e.g., JPEG, BPG) and channel coding (e.g., LDPC, Polar code) design, the original image undergoes feature extraction and compression through the JSC encoding at the transmitter, which is a non-linear mapping from the semantic information embedded in $\mathbf{x}$ into the k-dim complex-valued vector $\mathbf{s} \in \mathbb{C}^K$. This process can be expressed by

$$\mathbf{s} = f\left(\mathbf{x}; \boldsymbol{\theta}_{\mathrm{B}}\right), \tag{1}$$

where $\boldsymbol{\theta}_{\mathrm{B}}$ represents the trainable parameters in JSC encoder at the BS. Then, the signal goes through the attenuation of wireless channels, which will be discussed in Section II-B.

We denote the received complex-value signal at Bob and Eve as $\mathbf{r}_{\mathrm{L}}$ and $\mathbf{r}_{\mathrm{E}}$, respectively. After reshaping, we can obtain two real-value vectors which can be denoted as $\mathbf{v}_{\mathrm{L}} \in \mathbb{R}^{2\,K}$ and $\mathbf{v}_{\mathrm{E}} \in \mathbb{R}^{2\,K}$. The reshaping process can be represented as

$$\mathbf{v}_i = g_{\mathrm{F}}\left(\mathbf{r}_i\right), \tag{2}$$

where $i \in \{\mathrm{L}, \mathrm{E}\}$ is set to differentiate Bob and Eve.

At Bob, the reshaped vector $\mathbf{v}_{\mathrm{L}}$ is fed into the JSC decoder for joint channel and source decoding. The final output at Bob is a classification result y_{L}, which stands for the predicted class. At the same time, Eve can execute similar decoding process and obtain y_{E}. The decoding process at Bob and Eve can be expressed by

$$y_i = g\left(\mathbf{v}_i; \boldsymbol{\theta}_i\right), \tag{3}$$

where $i \in \{\mathrm{L}, \mathrm{E}\}$, and $\boldsymbol{\theta}_i$ denotes the trainable parameters in JSC decoder at Bob or Eve.

B. STAR-RIS-Assisted Communication Model

In our settings, the STAR-RIS works in MS protocol, where all elements are divided into N_{t} elements for transmitting-only and $N_{\mathrm{r}} = N - N_{\mathrm{t}}$ elements for reflecting-only. Accordingly, the STAR-RIS's TCV and RCV can be expressed by

$$\mathbf{q}_{\mathrm{t}} = \left(\mathrm{e}^{\mathrm{j}\phi_1^{\mathrm{t}}}, \ldots, \mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{t}}}^{\mathrm{t}}}\right)^H, \tag{4}$$

and

$$\mathbf{q}_{\mathrm{r}} = \left(\mathrm{e}^{\mathrm{j}\phi_1^{\mathrm{r}}}, \ldots, \mathrm{e}^{\mathrm{j}\phi_{N_{\mathrm{r}}}^{\mathrm{r}}}\right)^H, \tag{5}$$

respectively, and $\phi_{n_{\mathrm{t}}}^{\mathrm{t}}, \phi_{n_{\mathrm{r}}}^{\mathrm{r}} \in [0, 2\pi)$, $\forall n_{\mathrm{t}} \in \{1, \ldots, N_{\mathrm{t}}\}$, $\forall n_{\mathrm{r}} \in \{1, \ldots, N_{\mathrm{r}}\}$. The STAR-RIS's transmission- and reflection-coefficient matrix are $\boldsymbol{\Theta}_{\mathrm{t}} = \mathrm{Diag}(\mathbf{q}_{\mathrm{t}}^H)$ and $\boldsymbol{\Theta}_{\mathrm{r}} = \mathrm{Diag}(\mathbf{q}_{\mathrm{r}}^H)$.

We assume the BS is equipped with M antennas, while Bob and Eve are equipped with a single antenna each. We denote $\mathbf{H}_{\mathrm{BT}} \in \mathbb{C}^{N_{\mathrm{t}} \times M}$, $\mathbf{H}_{\mathrm{BR}} \in \mathbb{C}^{N_{\mathrm{r}} \times M}$ and $\mathbf{h}_{\mathrm{BE}}^H \in \mathbb{C}^{1 \times M}$ as the channel from the BS to the STAR-RIS's transmitting elements, the channel from the BS to the STAR-RIS's reflecting elements and the channel from the BS to Eve, respectively. The channel from the STAR-RIS's transmitting elements to Bob and the channel from the STAR-RIS's reflecting elements to Eve are denoted as $\mathbf{h}_{\mathrm{IL}}^H \in \mathbb{C}^{1 \times N_{\mathrm{t}}}$ and $\mathbf{h}_{\mathrm{IE}}^H \in \mathbb{C}^{1 \times N_{\mathrm{r}}}$, respectively.[1]

[1] We assume the STAR-RIS is used for the entire window, thus the direct link between the BS and Bob is not consided in this paper.

The transmitted signal at the BS can be expressed by

$$\tilde{\mathbf{s}} = \mathbf{w}_{\mathrm{p}} s, \tag{6}$$

where symbol $s \in \mathbb{C}$ represents the basic unit of data transmission in our system, characterized as a continuous complex value with normalized power. The beamforming vector at the BS is denoted as $\mathbf{w}_{\mathrm{p}}$ with the constraint that $\|\mathbf{w}_{\mathrm{p}}\|_2^2 \leq P_{\mathrm{BS}}$.

Bob's received signal can be represented as

$$r_{\mathrm{L}} = \mathbf{h}_{\mathrm{IL}}^H \boldsymbol{\Theta}_{\mathrm{t}} \mathbf{H}_{\mathrm{BT}} \tilde{\mathbf{s}} + z_{\mathrm{L}}, \tag{7}$$

where $z_{\mathrm{L}} \sim CN(0, \sigma_{\mathrm{L}}^2)$ is the additive white Gaussian noise (AWGN) with zero mean and variance of σ_{L}^2.

Similarly, the received signal at Eve can be expressed by

$$r_{\mathrm{E}} = \mathbf{h}_{\mathrm{BE}}^H \tilde{\mathbf{s}} + \mathbf{h}_{\mathrm{IE}}^H \boldsymbol{\Theta}_{\mathrm{r}} \mathbf{H}_{\mathrm{BR}} \tilde{\mathbf{s}} + z_{\mathrm{E}}, \tag{8}$$

where $z_{\mathrm{E}} \sim CN(0, \sigma_{\mathrm{E}}^2)$ is the AWGN with variance of σ_{E}^2.

III. Proposed Method

In the proposed STAR-RIS-assisted privacy-preserved SemCom system, the STAR-RIS plays a two-fold role. One is to enhance the signal transmission of SemCom between the BS and Bob, by jointly optimizing the STAR-RIS's TCV and the BS's beamforming vector. The other is to prevent privacy leakage to Eve, which is achieved by designing the STAR-RIS's RCV to form interference at Eve. The realization mechanisms of these two objectives are described in Section III-A and III-B, respectively.

A. SemCom Empowering Mechanism

To enhance the signal transmission of SemCom between the BS and Bob, we aim to maximize the received SNR at Bob, which can be expressed as

$$\mathcal{P}1: \max_{\mathbf{w}_{\mathrm{p}}, \boldsymbol{\Theta}_{\mathrm{t}}} \frac{\left|\mathbf{h}_{\mathrm{IL}}^H \boldsymbol{\Theta}_{\mathrm{t}} \mathbf{H}_{\mathrm{BT}} \mathbf{w}_{\mathrm{p}}\right|^2}{\sigma_{\mathrm{L}}^2} \tag{9a}$$

$$\text{s.t.} \quad \phi_{n_{\mathrm{t}}}^{\mathrm{t}} \in [0, 2\pi), \tag{9b}$$

$$\|\mathbf{w}_{\mathrm{p}}\|_2^2 \leq P_{\mathrm{BS}}. \tag{9c}$$

As demonstrated in [14], when $\boldsymbol{\Theta}_{\mathrm{t}}$ is fixed, we can directly obtain the optimal beamforming vector as

$$\mathbf{w}_{\mathrm{p}} = \sqrt{P_{\mathrm{BS}}} \frac{\left(\mathbf{h}_{\mathrm{IL}}^H \boldsymbol{\Theta}_{\mathrm{t}} \mathbf{H}_{\mathrm{BT}}\right)^H}{\|\mathbf{h}_{\mathrm{IL}}^H \boldsymbol{\Theta}_{\mathrm{t}} \mathbf{H}_{\mathrm{BT}}\|}. \tag{10}$$

When $\mathbf{w}_{\mathrm{p}}$ is fixed, we can first let $\mathbf{a} = \mathrm{Diag}(\mathbf{h}_{\mathrm{IL}}^H) \mathbf{H}_{\mathrm{BT}} \mathbf{w}_{\mathrm{p}}$ and receive its phases as $\arg(\mathbf{a})$. Next, we can obtain the STAR-RIS's optimal TCV as

$$\mathbf{q}_{\mathrm{t}} = \mathrm{e}^{\mathrm{j} \cdot \arg(\mathbf{a})}. \tag{11}$$

By alternately updating $\mathbf{w}_{\mathrm{p}}$ and $\mathbf{q}_{\mathrm{t}}$ by (10) and (11) respectively, until convergence or reaching a maximum number of iterations, we can obtain high-quality solutions $\mathbf{w}_{\mathrm{p}}^*$ and $\mathbf{q}_{\mathrm{t}}^*$.

B. Privacy Protection Mechanism

By designing the RCV, the received signal at the STAR-RIS can be adjusted and reflected to form interference at Eve. In this section, we propose two RCV designs, which target to generate task-level and SNR-level disturbance, respectively.

1) RCV Design for Generating Task-Level Interference: To degrade Eve's task performance, we generate adversarial signals which are characterized as subtly crafted imperceptible perturbations $\mathbf{r}_{\mathrm{A}} \in \mathbb{C}^K$ to Eve. The fusion of $\mathbf{r}_{\mathrm{A}}$ and $\mathbf{r}_{\mathrm{E}}$ can effectively prevent Eve's decoder from inferring task-related information embedded in $\mathbf{r}_{\mathrm{E}}$ and thus fool Eve to obtain a different task result (i.e. $g(g_{\mathrm{F}}(\mathbf{r}_{\mathrm{E}} + \mathbf{r}_{\mathrm{A}}); \boldsymbol{\theta}_{\mathrm{E}}) \neq g(g_{\mathrm{F}}(\mathbf{r}_{\mathrm{E}}); \boldsymbol{\theta}_{\mathrm{E}})$). In our system, $\mathbf{r}_{\mathrm{A}}$ denotes the transformed transmitted signal, resulting from the joint effect of wireless channels and STAR-RIS's RCV.

In adversarial learning, fast gradient sign method (FGSM) is used to generate adversarial examples for enhancing neural networks robustness, which is a onestep gradient-based method developed on finding the scaled sign of the gradient of the cost function and aims at minimizing the strength of the perturbation [15]. However, Bob's task performance can be affected if the adversarial examples are added to the request images directly [16]. Therefore, we only use FGSM to determine the adversarial signal's phases.

Based on the shared KBs, the FGSM-generated perturbation $\boldsymbol{\eta} \in \mathbb{R}^{2\,K}$ can be obtained by linearizing Eve's cost function $J(\theta_{\mathrm{E}}, \mathbf{v}_{\mathrm{E}}, y_{\mathrm{E}})$ around the current value of $\mathbf{v}_{\mathrm{E}}$:

$$\boldsymbol{\eta} = \mathrm{sign}(\nabla_{\mathbf{v}_{\mathrm{E}}} J\left(\theta_{\mathrm{E}}, \mathbf{v}_{\mathrm{E}}, y_{\mathrm{E}}\right)). \tag{12}$$

By inversely reshaping $\boldsymbol{\eta}$ to $\boldsymbol{\gamma}$ via $\boldsymbol{\gamma} = g_{\mathrm{F}}^{-1}(\boldsymbol{\eta})$, we can obtain the corresponding target phases for the complex-value adversarial signal $\mathbf{r}_{\mathrm{A}}$.

It has been shown that the reconfiguration time for the reconfigurable intelligent surface to change the response matrix is around 33 ns [17], which is capable of achieving real-time tuning [18]. By dynamically adjusting the phase shifts over time, the reflection elements on the STAR-RIS can be utilized to steer the phases of the complex-valued desired signal $\tilde{\mathbf{s}}$, transmitted from the BS towards the desired direction.

The phase-shift design for each reflection element n_{r} of STAR-RIS for each transmitted symbol s can be expressed by

$$\phi_{n_{\mathrm{r}}}^{\mathrm{r}} = \arg\left(\gamma\right) - \arg\left(\left(\mathrm{Diag}\left(\mathbf{h}_{\mathrm{IB}}^H\right) \mathbf{H}_{\mathrm{BR}} \mathbf{w}_{\mathrm{p}}\right)_{n_{\mathrm{r}}}\right) - \arg(s), \tag{13}$$

where $(\cdot)_{n_{\mathrm{r}}}$ denotes the n_{r}-th element of a vector. By substituting (13) into (5), we can obtain the RCV as $\mathbf{q}_{\mathrm{r}}^*$.

Thus, the interference signal at Eve can be written as

$$r_{\mathrm{A}} = \left(\mathbf{q}_{\mathrm{r}}^*\right)^H \mathrm{Diag}\left(\mathbf{h}_{\mathrm{IB}}^H\right) \mathbf{H}_{\mathrm{BR}} \mathbf{w}_{\mathrm{p}} s. \tag{14}$$

In this way, the interference is formed at Eve. Meanwhile, it causes no disturbance to the transmitted signal to Bob.

2) RCV Design for Generating SNR-Level Interference: For each received symbol at Eve, the SNR-level interference is generated to degrade its SNR, which can be formulated as:

$$\mathcal{P}2: \min_{\mathbf{q}_{\mathrm{r}}} \frac{\left|\mathbf{h}_{\mathrm{BE}}^H \mathbf{w}_{\mathrm{p}} + \mathbf{q}_{\mathrm{r}}^H \mathrm{Diag}\left(\mathbf{h}_{\mathrm{IB}}^H\right) \mathbf{H}_{\mathrm{BR}} \mathbf{w}_{\mathrm{p}}\right|^2}{\sigma_{\mathrm{E}}^2} \tag{15a}$$

$$\text{s.t.} \ \phi_{n_{\mathrm{r}}}^{\mathrm{r}} \in [0, 2\pi). \tag{15b}$$

To solve problem $\mathcal{P}2$, we can minimize the numerator by letting $\arg(\mathbf{q}_{\mathrm{r}}^H \mathrm{Diag}(\mathbf{h}_{\mathrm{IB}}^H) \mathbf{H}_{\mathrm{BR}} \mathbf{w}_{\mathrm{p}}) = -\arg(\mathbf{h}_{\mathrm{BE}}^H \mathbf{w}_{\mathrm{p}})$ to obtain $\mathbf{q}_{\mathrm{r}}^*$. If the power of $|\mathbf{q}_{\mathrm{r}}^H \mathrm{Diag}(\mathbf{h}_{\mathrm{IB}}^H) \mathbf{H}_{\mathrm{BR}} \mathbf{w}_{\mathrm{p}}|^2$ exceeds $|\mathbf{h}_{\mathrm{BE}}^H \mathbf{w}_{\mathrm{p}}|^2$, some STAR-RIS's reflection elements can be tuned off to ensure the received signal's power at Eve is the lowest.

IV. Simulation Results

In this section, we conduct a series of experiments to evaluate the performance of the proposed schemes with other benchmark methods.

Dataset: We use MNIST [19] for training and testing. It contains 60,000 training images and 10,000 testing images. The image's dimension is $28 \times 28 \times 1$. The whole image dataset is composed of 10 classes.

Channel Settings: We assume that the locations of the BS, STAR-RIS, Eve and Bob are (0 m, 0 m), (Lm, 5 m), (40 m, 0 m) and (100 m, 10 m), where L is equal to 40 by default. We set $\sigma_{\mathrm{L}}^2 = \sigma_{\mathrm{E}}^2 = -90\,\mathrm{dBm}$ [20]. The large-scale fading is modelled by $\mathrm{PL}(d) = \mathrm{PL}_0(d/d_0)^{\varpi}$, where $\mathrm{PL}_0 = 30\,\mathrm{dB}$ is the path loss at the reference distance $d_0 = 1\mathrm{m}$, d is the distance, and ϖ is the path-loss exponent which is set to 2.5 [21]. For small-scale fading, the Rayleigh fading model is assumed for all channels [21]. Because of difference in distance and "multiplicative fading" effect [22], it can be verified that Eve's channel condition is better than Bob in our settings.

Training Settings: We first train the JSC encoding and decoding networks for SemCom between the BS and Bob, with the BS's beamforming vector and STAR-RIS's TCV designed as Section III-A. To demonstrate the effectiveness of the proposed methods, a worst-case privacy-leakage scenario is considered, where the downstream task is the same classification problem at both receivers and the decoder at Eve is trained to achieve high privacy eavesdropping performance before the interference is generated. It should be noted that for fair comparison, the parameters of Bob's or Eve's decoders are the same for all methods and the task success rate at Bob are trained to be equal. The cross entropy is used as our loss function. We fine-tune the encoder's and decoders' networks based on pre-trained LeNet [19]. The Adam optimizer is adoped with a learning rate of 0.01. The experiments are implemented on a NVIDIA GTX 1080 Ti GPU.

Benchmark Methods: We compare the proposed task-level protection (TLP) method and SNR-level protection (SNRLP) method with the following benchmarks:

- Task-level protection with random reflection phases (TLP-random): STAR-RIS is still deployed and its TCV is optimized using our proposed method while its reflection phase shifts are randomly selected in $[0, 2\pi)$.
- JSC SemCom without protection (Without Protection): STAR-RIS is still deployed in the system to assist the semantic communication while no privacy protection methods are adopted to prevent information leakage.
- Secure JSC autoencoder design with training (SET) [11]: The SemCom is protected by training an autoencoder at the BS. The criterion of privacy leakage is modified in order to suit our classification task. Specifically, the whole loss function balances both efficiency and privacy. The loss for protecting privacy can be expressed by:

$$l'_{\mathrm{E}}(k) = \begin{cases} -l_{\mathrm{E}}(k), & \text{if softmax}\left(\mathbf{y}_{\mathrm{E},k}^{\mathrm{P}}\,(\mathrm{target})\right) > \varepsilon, \\ 0, & \text{otherwise}, \end{cases} \tag{16}$$

where k denotes the k-th sample in a batch, $\mathbf{y}_{\mathrm{E},k}^{\mathrm{P}}(\mathrm{target})$ denotes the target index of the predict vector of Eve's decoder network, ε is a predefined indicator of privacy eavesdropping. Then, the loss function with privacy-aware for training the encoder at the BS is given by

$$l_{\mathrm{total}} = \frac{1}{B}\sum_{k=1}^{B}\left[l_{\mathrm{L}}(k) - \lambda \cdot l'_{\mathrm{E}}(k)\right], \tag{17}$$

where $l_{\mathrm{L}}(k)$ is sample k's loss function at Eve, B is the number of samples in a batch, λ is the weighting factor.

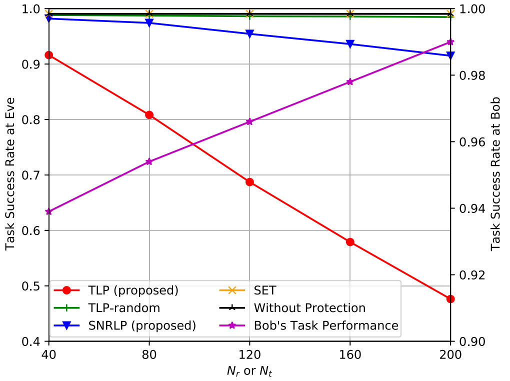


Fig. 3. Task success rate versus N_{r} or N_{t}.

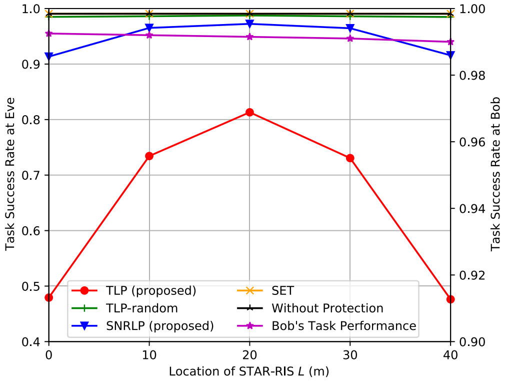
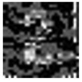
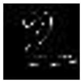
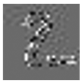
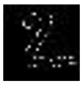
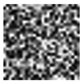
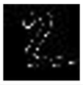
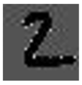
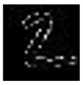
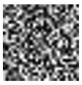
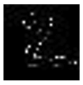


Fig. 4. Task success rate versus location of STAR-RIS L.

We first present the comparison of TLP, SNRLP and TLP-random in Table I, where the attack and fusion signals at Eve are transformed to the same type as the original MNIST images. The compressed rate equals 1 for better visualization. The results show that our proposed TLP generates the strongest interference to Eve, and the privacy of SemCom between the BS and Bob is significantly enhanced. From the attack designs, it can be observed that the attack performed by the TLP-random manifests as random noise, which achieves the weakest disturbance to Eve. Furthermore, the results imply that minimizing the SNR at Eve is not the optimal solution to protect privacy in task-oriented SemCom. In contrast, our proposed TLP strategically designs the attack. Aligning with the objective of increasing the loss function at Eve, the perturbed input causes Eve's model to output an incorrect answer with high confidence. Besides, with the increase in STAR-RIS's reflection element number N_{r}, the strength of the attack becomes stronger, leading to more severe degradation in Eve's task success rate.

Eve's and Bob's task success rates versus STAR-RIS's reflection element number N_{r} and transmission element number N_{t} are presented in Fig. 3. The performance of the systems without STAR-RIS, namely "SET" and "Without Protection", remains unaffected by the variation in N_{r}. The SET fails to take effect in our system, which is because that the SET requires Eve's channel condition to be worse than Bob's. However, due to the ease of eavesdropping on data, in practice, Eve tends to choose locations with better channel conditions. In our settings, the TLP-random exhibits slightly better performance compared to the system without protection, where the success rate at Eve is still above 90% in our settings. Eve's performance in our proposed TLP and SNRLP is noticeably degraded with the increase of N_{r}, which is due to the fact that a larger number of reflection elements can contribute to more attack energy concentrated at Eve. Specifically, our proposed

TABLE I
VISUALIZATION COMPARISON OF TLP, SNRLP AND TLP-RANDOM

N_r	Method	Original Image	Attack Design	Image Received at Eve	Task Success Rate at Eve
40	TLP				91.62%
	SNRLP				98.20%
	TLP-random				98.51%
200	TLP				47.62%
	SNRLP				91.50%
	TLP-random				98.48%

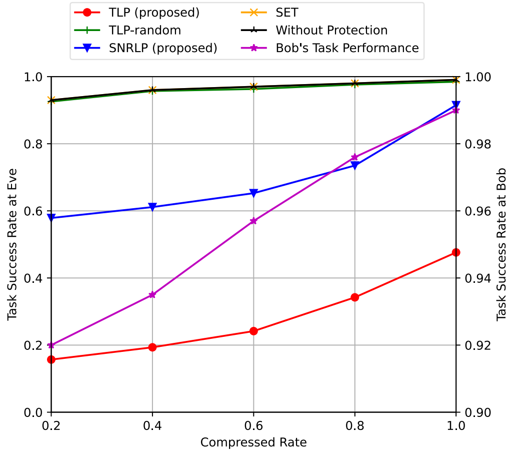


Fig. 5. Task success rate versus compressed rate.

TLP outperforms SNRLP, highlighting the advantages of generating task-oriented interference for privacy protection over solely reducing the SNR at Eve in SemCom. Meanwhile, the increase in N_t boosts Bob's task performance, showcasing the STAR-RIS's dual functionality in enhancing SemCom between the BS and Bob as well as improving privacy.

To underscore the efficacy of our proposed schemes in relation to the varying positions of STAR-RIS, we present an analysis of Bob and Eve's task performance in relation to the STAR-RIS's location L in Fig. 4. Our findings reveal that situating the STAR-RIS nearer to the BS results in a slight improvement in Bob's task performance. Additionally, while our protection schemes show superior performance compared to other benchmarks, it is noteworthy that the performance of Eve exhibits symmetry around the center of the BS and Eve's position. Importantly, the success rate of Eve is lower when the STAR-RIS is positioned closer to either the BS or Eve. This is because in the range of $L \in [0, 40]$, the path loss of cascaded BS-(STAR-RIS)-Eve channel reaches the maximum at $L = 20$ and decreases symmetrically from the center. Drawing from these observations, it is advisable to position the STAR-RIS closer to the BS in practical applications.

We also explore the influence of the compressed rate on both Bob's and Eve's task performance, as shown in Fig. 5. These results further emphasize the superiority of our proposed method in achieving privacy protection in SemCom. As depicted in Fig. 5, all methods exhibit stronger interference to Eve as the compressed rate decreases. This observation suggests that when the transmitted information is highly compressed, it becomes more susceptible to attacks. It is also important to note that the task success rate at Bob also decreases with the compressed rate. This is because the compression process eliminates less important content for the task, thereby reducing the redundancy and weakening the error correction capacity of the transmitted information itself.

V. CONCLUSION

In this paper, we have presented a STAR-RIS-assisted privacy protection system to achieve two main objectives. First, we have enhanced the SemCom between the BS and Bob by jointly optimizing the TCV of the STAR-RIS and the beamforming vector of the BS. Second, we have addressed the privacy leakage by designing the RCV of the STAR-RIS to create task-level and SNR-level interference for Eve, respectively. By converting the desired signal into interference for Eve, we have degraded its task performance while causing no disturbance to the SemCom between the BS and Bob. The results of our simulations have highlighted the advantages of generating task-related interference over SNR-level disturbance in protecting the privacy of SemCom. Furthermore, we have demonstrate the proposed scheme's superiority over existing methods, especially when Eve's channel condition is better than that of Bob. Our study's main limitation lies in the reliance on specific channel expressions. Moreover, advanced defense systems might detect and counter our method. Nonetheless, our work serves as a foundational stepping stone for future research.

REFERENCES

[1] W. Yang et al., "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 213–250, Firstquarter 2023.

[2] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Commun.*, vol. 29, no. 1, pp. 210–219, Feb. 2022.

[3] E. Tekin and A. Yener, “The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, Jun. 2008.

[4] S. Goel and R. Negi, “Guaranteeing secrecy using artificial noise,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.

[5] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, “Deep joint source-channel coding for wireless image transmission,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2019, pp. 4774–4778.

[6] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, “Deep learning enabled semantic communication systems,” *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2023.

[7] D. Adesina, C.-C. Hsieh, Y. E. Sagduyu, and L. Qian, “Adversarial machine learning in wireless communications using RF data: A review,” *IEEE Commun. Surveys Tuts.*, vol. 25, no. 1, pp. 77–100, Firstquarter, 2023.

[8] M. Sadeghi and E. G. Larsson, “Adversarial attacks on deep-learning based radio signal classification,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 213–216, Feb. 2019.

[9] T. Hou et al., “MUSTER: Subverting user selection in MU-MIMO networks,” in *Proc. IEEE Conf. Comput. Commun.*, 2022, pp. 140–149.

[10] Y. E. Sagduyu, Y. Shi, and T. Erpek, “Adversarial deep learning for over-the-air spectrum poisoning attacks,” *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 306–319, Feb. 2021.

[11] M. Zhang, Y. Li, Z. Zhang, G. Zhu, and C. Zhong, “Wireless image transmission with semantic and security awareness,” *IEEE Wireless Commun. Lett.*, vol. 12, no. 8, pp. 1389–1393, Aug. 2023.

[12] Y. Liu et al., “STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces,” *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, Dec. 2021.

[13] J. Zhu, P. Gao, G. Chen, P. Xiao, and A. Quddus, “Index modulation for STAR-RIS assisted NOMA system,” *IEEE Commun. Lett.*, vol. 27, no. 2, pp. 716–720, Feb. 2023.

[14] Q. Wu and R. Zhang, “Intelligent reflecting surface enhanced wireless network: Joint active and passive beamforming design,” in *Proc. IEEE Glob. Commun. Conf.*, 2018, pp. 1–6.

[15] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” Dec. 2014, *arXiv:1412.6572*.

[16] Y. Wen, S. Li, and K. Jia, “Towards understanding the regularization of adversarial robustness on neural networks,” in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 10225–10235.

[17] T. J. Cui et al., “Information metamaterial systems,” *iScience*, vol. 23, no. 8, Aug. 2020, Art. no. 101403.

[18] H. Du et al., “Semantic communications for wireless sensing: RIS-aided encoding and self-supervised decoding,” *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2547–2562, Aug. 2023.

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[20] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, “Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, pp. 3083–3098, May 2022.

[21] P. Guan, Y. Wang, H. Yu, and Y. Zhao, “Joint beamforming optimization for RIS-aided full-duplex communication,” *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1629–1633, Aug. 2022.

[22] Z. Zhang et al., “Active RIS vs. passive RIS: Which will prevail in 6G?,” *IEEE Trans. Commun.*, vol. 71, no. 3, pp. 1707–1725, Mar. 2023.