Optimization of Private Semantic Communication Performance: An Uncooperative Covert Communication Method

Wenjing Zhang, *Student Member, IEEE*, Ye Hu, *Member, IEEE*, Tao Luo, *Senior Member, IEEE*, Zhilong Zhang, *Member, IEEE*, and Mingzhe Chen, *Senior Member, IEEE*

***Abstract*—In this paper, a novel covert semantic communication framework is investigated. Within this framework, a server extracts and transmits the semantic information, i.e., the meaning of image data, to a user over several time slots. An attacker seeks to detect and eavesdrop the semantic transmission to acquire details of the original image. To avoid data meaning being eavesdropped by an attacker, a friendly jammer is deployed to transmit jamming signals to interfere the attacker so as to hide the transmitted semantic information. Meanwhile, the server will strategically select time slots for semantic information transmission. Due to limited energy, the jammer will not communicate with the server and hence the server does not know the transmit power of the jammer.Therefore, the server must jointly optimize the semantic information transmitted at each time slot and the corresponding transmit power to maximize the privacy and the semantic information transmission quality of the user. To solve this problem, we propose a prioritised sampling assisted twin delayed deep deterministic policy gradient algorithm to jointly determine the transmitted semantic information and the transmit power per time slot without the communications between the server and the jammer. Compared to standard reinforcement learning methods, the proposed method uses an additional Q network to estimate Q values such that the agent can select the action with a lower Q value from the two Q networks thus avoiding local optimal action selection and estimation bias of Q values. Simulation results show that the proposed algorithm can improve the privacy and the semantic information transmission quality by up to 77.8% and 14.3% compared to the traditional reinforcement learning methods.**

***Index Terms*—Semantic communication, covert communication, reinforcement learning.**

I. Introduction

Current communication techniques (e.g., reflected intelligent surface [1] and non-terrestrial communications [2]) may not be able to support emerging wireless applications, especially those AI-enabled services, e.g., automatic driving, digital twins, and Metaverse, that require to reliably and efficiently transmit massive volumes of image data that collected by dense visual devices [3]–[5]. Semantic communication [6]–[10] is a novel and promising paradigm to support these resource-intensive services [11], by focusing on transmitting only the most relevant meaning of the original data (called *semantic information*) in the receiver's need [12]. However, in a semantic communication system, the transmitted semantic information that extracted and refined from the original data based on the AI-enabled encoder is much more meaningful [13]–[15], sensitive, and private. A malicious attacker can acquire more valuable information by eavesdropping semantic information in a semantic communication system compared to that in standard communication systems [16]. Consequently, the security or privacy issues in semantic communication become more critical [17]. Covert communication techniques that can hide the very existence of transmission from a malicious attacker are considered as a powerful solution for transmitting massive image data efficiently and securely. However, introducing covert transmission techniques into semantic communication faces several challenges including semantic-related transmit power control, transmission time slot arrangement, evaluation of the semantic information quality, and privacy of the semantic communication system.

A. Related Works

Recently, several works [18]–[25] investigated the security issues in semantic communication systems. The authors in [18] combined semantic coding and digital watermark technique to protect the semantic of the transmitted data from abusing and tampering. The authors in [19] developed a secure semantic communications framework to prevent data positioning attacks via using blockchains and zero-knowledge proofs. The authors in [20] introduced an adversarial training method to against various semantic-oriented physical adversarial attacks from an imperceptible physical-layer adversarial perturbation generator in semantic transmission. The authors in [21] applied diffusion model to defend semantic-oriented attacks by adversarial purification. However, these works [18]–[21] do not consider eavesdropping attack, which also poses significant security threats to semantic communications systems. The works in [22] introduced physical layer encryption and subcarrier obfuscation solutions to support the security of semantic communication from eavesdropping attack. The work in [23] introduced an unified secure semantic communication framework with three hot-plug-gable semantic protection modules that can secure the transmission by encryption, mitigate leakage risk by perturbation, and calibrate distortion in the receiver by semantic signature generation. The work in [24] designed a random permutation and substitution method to prevent the

W. Zhang, T. Luo, and Z. Zhang are with the Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications, Beijing, 100876, China (e-mail zhangwenjing@bupt.edu.cn; tluo@bupt.edu.cn; zhangzhilong@bupt.edu.cn).

Y. Hu is with the Department of Industrial and Systems Engineering, University of Miami, Coral Gables, FL, 33146 USA (Email: yehu@miami.edu).

M. Chen is with the Department of Electrical and Computer Engineering and Institute for Data Science and Computing, University of Miami, Coral Gables, FL, 33146 USA (Email: mingzhe.chen@miami.edu).

attacker decoding the eavesdropped semantic information. In [25], the authors introduced a privacy aware loss function that guide the training of joint source and channel coding based neural networks so as to improve the image reconstruction quality while reducing data leakage. However, these encryption and perturbation based methods [22]–[25] focused on the protection of the transmitted information and used the size of protected data as a metric to evaluate protection performance. Hence, they do not consider data meaning eavesdropped by the attacker. In practice, an attacker may not be interested about the entire transmitted data. Hence, considering the data meanings that the attacker is interested can significantly improve the data privacy performance.

Covert communications is a potential solution to secure the wireless transmission with high-security level [26]. Compared to the traditional physical layer security, covert communication technique can hide the very existence of the wireless transmission without the spectral cost of tedious and redundant encryption management process, by properly controlling transmit power [27] or interfering malicious detection with the help of a friendly jammer [28]. Hence, some works in [29]–[32] studied the use of covert communication technique in semantic communication system to improve semantic security. In [29], a covert and reliable semantic communication framework is introduced to against multiple eavesdropping attacks by using a full-duplex receiver that is required to decode the received semantic information and transmit artificial noise simultaneously. Similarly, in [30], the full-duplex receiver is also considered to construct a covert semantic communication system. However, these works [29], [30] may not be used for energy limited devices since they require to use full-duplex receiver and self-interference concealment schemes, which are energy consuming. In [31], a multi-agent reinforcement learning algorithm is proposed to secure the semantic communication from eavesdropping attack by selectively, and thus temporally, protecting one user with the friendly jammer in the system. Hence, all the rest of users are exposed to the attack's eavesdropping without any protection. In [32], a generative diffusion model based covert communication method is developed to hide the semantic transmission. Despite the promising results, these two works [31], [32] on covert semantic communication require the connection and cooperation between the transmitter and the jammer, which is cost-expensive and impractical for the most wireless communication scenarios due to the limited energy and computation capacity of the jammer.

B. Contributions

The main contribution of this work is a novel covert semantic communication framework that enables the server to consistently and privately transmit semantic information to the user against eavesdropping attacks without information sharing with a friendly jammer. The key contributions include:

- We propose a novel covert semantic communication framework within which a server transmits image data to a user utilizing semantic communication techniques, while a friendly jammer is deployed to protect this user from multiple eavesdropping attacks during semantic transmission duration. The server and friendly jammer cannot communicate with each other such that the server transmitting semantic information without the help of channel state information and transmit power of the jammer. Therefore, the jammer in the proposed framework is also not required to estimate wireless channel and coordinate with the server in choosing jamming signal power. To improve the quality of the received semantic information of the user and the system privacy, the server must jointly optimize the partial semantic information transmitted and the corresponding transmit power at each transmission time slot without aid of connection and coordination from the jammer.
- To evaluate the joint semantic communication and privacy performance of the semantic transmission, a novel metric called graph-to-nearest-triple (GNT) is introduced. This metric can directly capture the correlation of the meanings of the extracted and received semantic triples, considering the image contents rather than bit errors, and does not require access to the embedding of the original image.
- We formulate this semantic triple selection and transmit power control problem as an optimization problem whose goal is to maximize the semantic similarity of the user while guaranteeing the semantic similarity of the attacker being low. Since the optimization problem is non-convex and the objective function and semantic communication

TABLE I: List of notations

Notation	Description
n	Index of transmission time slot
N	Number of transmission time slots
B	Number of triple transmitted at one time slot
W	Bandwidth for transmission
p_J^n	Transmit power of the jammer
p_S^n	Transmit power of the server
β	Path loss exponent
σ_U	Standard deviation of Gaussian noise
$\boldsymbol{q}_n$	triple transmission vector
ζ_U^n	Power of the received signal
c_U^n	Downlink data rate
t_U^n	Transmission latency
T	Transmission latency threshold
$\boldsymbol{\Psi}$	Semantic information
$\boldsymbol{\psi}^k$	Semantic triple k in $\boldsymbol{\Psi}$
$Z\left(\boldsymbol{\psi}^k\right)$	Number of bits to transmit $\boldsymbol{\psi}^k$
$\boldsymbol{g}$	Transmission detection vector
ϵ_n	Power detection threshold
$C\left(\boldsymbol{\psi}^k\right)$	Vectorized semantic triple $\boldsymbol{\psi}^k$
E_U	Semantic similarity metric
G	Maximum detection number
γ	Semantic privacy threshold
η	Penalty of insecure transmission

privacy constraint depends on graph embedding neural network model, the relationship between the objective function and optimization variables cannot be represented exactly. In consequence, such a maximization problem cannot be solved by traditional optimization methods. To solve the problem, we introduce a prioritized sampling assisted deep RL method that can be trained offline with less interactions between the server and the user such that reducing experience collection overhead.

- To prevent the RL algorithm overestimating Q values thus converging to sub-optimal policies, we introduce a clipped double Q learning technique that can learn Q value function in a conservative and steady way. Meanwhile, to improve the convergence speed of this delayed RL, we also introduce a prioritized sampling mechanism that enables the agent to learn from more valuable experience by including the sampling priority in historical experience replaying in training stage.

The rest of this paper is organized as follows. The proposed covert image semantic communication system model and the problem formulation are described in Section II. Section III introduces the proposed prioritized sampling assisted twin delayed deep deterministic policy gradient algorithm for private semantic transmission optimization. In Section IV, numerical results are presented and discussed. Finally, conclusion are drawn in Section V.

II. System Model and Problem Formulation

Consider a cellular network in which a server transmits the meaning of images to a user using semantic communication techniques while avoiding detection by an eavesdropping attacker. In the considered model, as shown in Fig. 1, the server will determine whether to transmit semantic information at each time slot. We assume that the server needs to use N time slots to complete semantic information transmission. The attacker can only select G time slots from N time slots to eavesdrop data due to its limited energy. To achieve covert semantic communications, a friendly jammer is used to protect the semantic information transmitted by the server via transmitting interference jamming signals. Here, the jammer cannot communicate with the server or the user such that the jammer does not know the transmit power of the server while the server does not know the power of the attacker transmitting a jamming signal at each time slot. Next, we first introduce the procedure of the semantic information extraction. Then, we present the covert semantic information transmission. Finally, we define a metric to evaluate the quality of the received semantic information and introduce the problem formulation. Table I summarizes all parameters used in our work.

A. Semantic Information Extraction

We model the semantic information of each image by a scene graph that consists of several nodes and edges, where a node represents an object and an edge represents the relationship between two objects. Hence, the semantic information consists of both the objects and their relationships in the image. We define a basic component of semantic information, called semantic triple. Each triple consists of two objects and the relationship between them. For example, as shown in Fig. 2, a semantic triple is ([“*man*”], [“*holding*”], [“*bag*”]), where [“*holding*”] is the relationship between objects [“*man*”] and [“*bag*”]. Therefore, a server only needs to transmit multiple semantic triples that can describe the original image accurately.

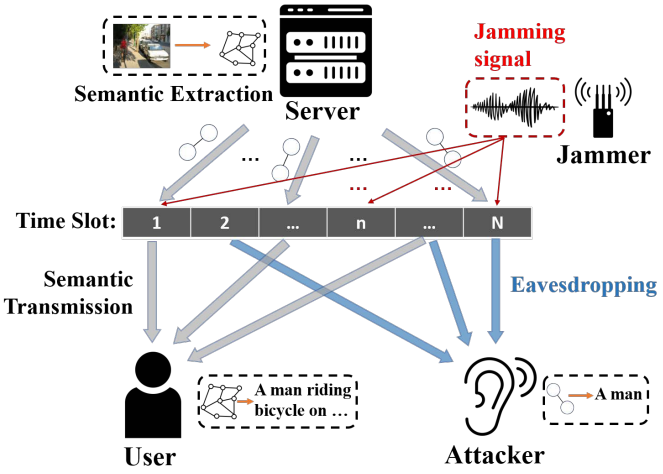


Fig. 1. Secure semantic communication wireless network.

To obtain the semantic information from the original image, a scene graph extraction model in [33] is used. Specifically, the server firstly uses the model to detect, locate, and categorize all the objects in the original image. Then, the relationships between all the objects are deduced and represented in form of triples. The extracted semantic information is

$$\boldsymbol{\Psi} = \left\{\boldsymbol{\psi}^1, \boldsymbol{\psi}^2, \ldots, \boldsymbol{\psi}^k, \ldots, \boldsymbol{\psi}^K\right\}, \tag{1}$$

where K is the number of semantic triples in image and $\boldsymbol{\psi}^k = \left(e_k^i, l_k, e_k^j\right)$ is a semantic triple with l_k being the relationship between objects e_k^i and e_k^j.

B. Covert Semantic Transmission

We assume that the server transmits K semantic triples extracted from the original image via N transmission time slots. Specifically, a triple transmission vector over N time slots can be given by

$$\boldsymbol{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n, \ldots, \boldsymbol{q}_N], \tag{2}$$

where $\boldsymbol{q}_n = [q_1^n, q_2^n, \ldots, q_k^n, \ldots, q_K^n]$ is triple transmission vector at time slot n with $q_k^n = 1$ representing that the server will transmit semantic triple ψ^k at slot n, and $q_k^n = 0$, otherwise. We assume that the semantic triples are transmitted over AWGN channels. Then, the power of the received signal at the user ζ_U^n and attacker ζ_A^n at time slot n can be given as

$$\zeta_U^n = \frac{p_S^n}{d_{S,U}^{-\beta}} + \frac{p_J^n}{d_{J,U}^{-\beta}} + \sigma_U^2 \tag{3}$$

and

$$\zeta_A^n = \frac{p_S^n}{d_{S,A}^{-\beta}} + \frac{p_J^n}{d_{J,A}^{-\beta}} + \sigma_A^2, \tag{4}$$

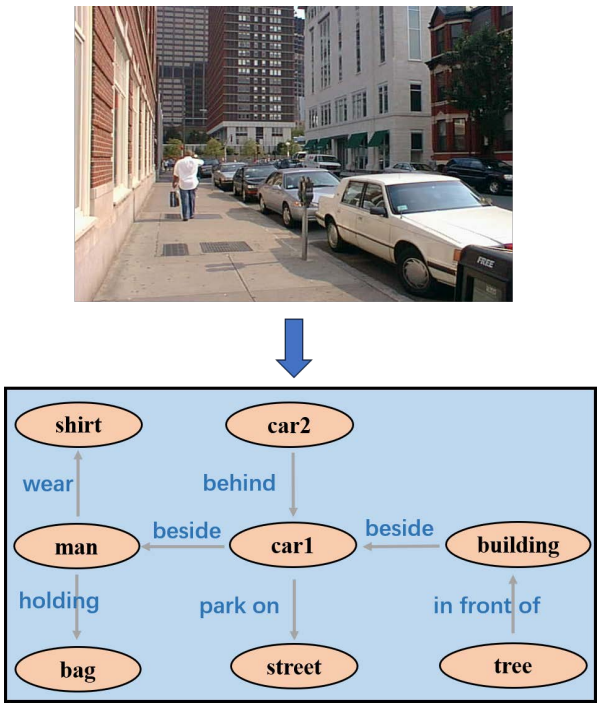


Fig. 2. Semantic information extraction.

where p_S^n and p_J^n are the transmit power of the server and the jammer at time slot n, respectively, $d_{S,U}$ and $d_{S,A}$ are the distance from the server to the user and the attacker, $d_{J,U}$ and $d_{J,A}$ are the distance from the jammer to the user and the attacker, β is the path loss exponent, σ_U^n and σ_A^n are the variance of the Gaussian noise at the user and the attacker, respectively. The transmit power p_J^n of the jammer at each time slot n follows an uniformed distribution with a range $[0, P_J^{\max}]$.

The data rate of the server transmitting a semantic triple to the user at time slot n can be given as

$$c_U^n = W \log\left(1 + \frac{p_S^n d_{S,U}^{-\beta}}{p_J^n d_{J,U}^{-\beta} + \sigma_U^2}\right), \tag{5}$$

where W is the bandwidth. Then, the transmission latency of semantic triple $\boldsymbol{\psi}^k$ at time slot n is $t_U^n = \frac{Z(\boldsymbol{\psi}^k)}{c_U^n}$, where function $Z\left(\boldsymbol{\psi}^k\right)$ is the number of bits that the server requires to transmit $\boldsymbol{\psi}^k$ over wireless links. Similarly, we can obtain the transmission data rate and latency of the attacker as c_A^n and t_A^n.

In the considered system, the semantic transmission fails and the received incomplete semantic triple will be discarded when transmission latency is larger than the predefined latency threshold T. The user will not ask the server to resend the discarded semantic information. Therefore, the received semantic information at the user is

$$\boldsymbol{\Psi}_U\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}\right) = \left\{\boldsymbol{\psi}_U^1, \boldsymbol{\psi}_U^2, \ldots, \boldsymbol{\psi}_U^k, \ldots, \boldsymbol{\psi}_U^K\right\}, \tag{6}$$

where $\boldsymbol{\psi}_U^k$ is the received semantic triple that is given by

$$\boldsymbol{\psi}_U^k = \begin{cases} \boldsymbol{\psi}^k, & t_U^n \leqslant T, \\ \mathbf{0}, & t_U^n > T, \end{cases} \tag{7}$$

$\boldsymbol{p}_S = \left[p_S^1, \ldots, p_S^n, \ldots, p_S^N\right]$, and $\boldsymbol{p}_J = \left[p_J^1, \ldots, p_J^n, \ldots, p_J^N\right]$ are transmit power vector of the server and the jammer. From (7), we see that if the transmission latency t_U^n is larger than the latency threshold T, the received semantic information will be null.

C. Detection and Eavesdropping of The Attacker

The attacker randomly selects a number of time slots to detect and eavesdrop due to its energy limitation. Here, we represent the detection of the server by a vector $\boldsymbol{g} = [g_1, g_2, \ldots, g_n, \ldots, g_N]$, where $g_n = 1$ represents that the attacker detects at time slot n, and $g_n = 0$, otherwise. Then, the attacker will determine if there is a semantic triple transmitted over a wireless channel at detected slot n. Here, a radiometer method [34], [35] is used for semantic triple transmission detection. In particular, the attacker compares the total received power with a power threshold ϵ_n. The detection result at time slot n is

$$D_n = \begin{cases} 1, & \zeta_A^n \geqslant \epsilon_n, \\ 0, & \zeta_A^n < \epsilon_n. \end{cases} \tag{8}$$

When $D_n = 1$, the attacker will decode the received signal to acquire transmitted semantic triple. From (8), we see that the attacker will not decode the received signal if its power is lower than the power threshold, i.e, $\zeta_A^n < \epsilon_n$. Meanwhile, the attacker also cannot decode a semantic triple if the received signal power is larger than the power threshold while the server does not transmit any triples. The semantic information eavesdropped by the attack is

$$\boldsymbol{\Psi}_A\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}, \boldsymbol{g}\right) = \left\{\boldsymbol{\psi}_A^1, \boldsymbol{\psi}_A^2, \ldots, \boldsymbol{\psi}_A^k, \ldots, \boldsymbol{\psi}_A^K\right\}, \tag{9}$$

where $\boldsymbol{\psi}_A^k$ is the eavesdropped semantic triple that is given by

$$\boldsymbol{\psi}_A^k = \begin{cases} \boldsymbol{\psi}^k, & q_k^n g_n D_n \mathbb{1}_{\left\{t_A^n \leqslant T\right\}} = 1, \\ \mathbf{0}, & \text{otherwise}. \end{cases} \tag{10}$$

From (10), we can see that the attacker successfully eavesdrops the transmission of the triple $\boldsymbol{\psi}_k$ at time slot n ($q_k^n = 1$) only when 1) the attacker selects time slot n to detect (i.e., $g_n = 1$), 2) the attacker successfully decode the semantic triple (i.e., $D_n = 1$), 3) the transmission latency t_A^n from the server to the attacker is smaller than the latency threshold T (i.e., $\mathbb{1}_{\left\{t_A^n \leqslant T\right\}} = 1$). Therefore, to prevent the attacker from eavesdropping the transmitted semantic information, the server have to jointly optimize transmit power and select triples to transmit at each time slot.

D. Metric for Semantic Communication

To evaluate the quality of the received semantic information and the eavesdropped semantic information, we cannot use mean square errors between the original image and the received data since there is no image reconstruction in the considered system. To this end, we introduce a metric called graph-to-nearest-triple (GNT). Different from the metric in [36] that measures the similarity between the extracted semantic information and the original image, the proposed GNT can capture the correlation of the meanings of the extracted

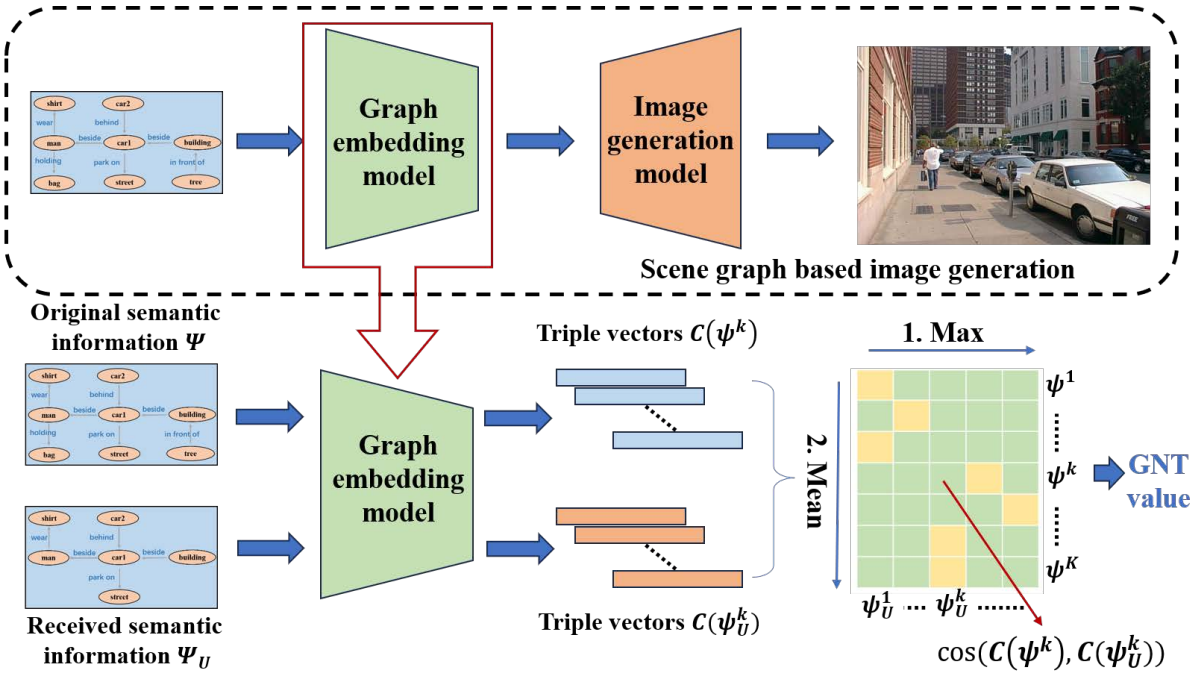


Fig. 3. Image-to-nearest-triple metric for semantic information.

and received semantic triples by directly comparing the image contents, thus, exclude the need for accessing the embedding of the original image.

The proposed metric must first transform the semantic information to a set of vectors by using a pre-trained scene graph embedding model $C(\cdot)$ as vectorization function. Then, based on this vectorization function, we compute cosine similarities between the original semantic triples and the received triples of user. Specifically, the cosine similarity between the original triple $\boldsymbol{\psi}^k$ and the received triple ψ_U^k is

$$\cos\left(C\left(\boldsymbol{\psi}^k\right), C\left(\boldsymbol{\psi}_U^k\right)\right) = \frac{C\left(\boldsymbol{\psi}^k\right) C\left(\boldsymbol{\psi}_U^k\right)^T}{\|C\left(\boldsymbol{\psi}^k\right)\| \|C\left(\boldsymbol{\psi}_U^k\right)\|}, \tag{11}$$

where $C\left(\boldsymbol{\psi}^k\right)$ is the semantic triple vector. Here, we note that $C\left(\boldsymbol{\psi}_U^k\right) = \mathbf{0}$ if $\boldsymbol{\psi}_U^k$ is null, which implies that this triple is not received by user k due to large transmission latency. For each triple in the original semantic information, we find its closest triple in the received semantic information that maximizes the value of the cosine similarity. Finally, the GNT is defined as

$$E_U\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}\right) = \frac{1}{K}\sum_{k=1}^{K} \max_{\boldsymbol{\psi}_U^k} \left\{\cos\left(C\left(\boldsymbol{\psi}^k\right), C\left(\boldsymbol{\psi}_U^k\right)\right)\right\}. \tag{12}$$

From (11) and (12), we see that the value of the GNT metric will be high if the user can always find a semantic triples $\boldsymbol{\psi}_U^k$ with similar meaning of triple $\boldsymbol{\psi}^k$.

Similarly, the GNT of the eavesdropped semantic information $\boldsymbol{\Psi}_A\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}, \boldsymbol{g}\right)$ is

$$E_A\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}, \boldsymbol{g}\right) = \frac{1}{K}\sum_{k=1}^{K} \max_{\boldsymbol{\psi}_A^k} \left\{\cos\left(C\left(\boldsymbol{\psi}^k\right), C\left(\boldsymbol{\psi}_A^k\right)\right)\right\}. \tag{13}$$

E. Problem Formulation

Given the defined system model, our objective is to maximize the semantic similarity of the user while ensuring the semantic similarity of the attacker being low. This maximization problem optimizes the triple transmission matrix $\boldsymbol{Q}$ and the transmit power vector $\boldsymbol{p}_S$ of the server, which is formulated as

$$\max_{\boldsymbol{p}_S, \boldsymbol{Q}} E_U\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}\right) - E_A\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}, \boldsymbol{g}\right) \tag{14}$$

$$\text{s.t.} \quad p_S^n \leqslant P_S^{\max}, \tag{14a}$$

$$p_J^n \sim U(0, P_J^{\max}), \tag{14b}$$

$$\sum_{k=1}^{K} q_k^n \leqslant B, \sum_{n=1}^{N} q_k^n = 1, \tag{14c}$$

$$\sum_{n=1}^{N} g_n \leqslant G, \tag{14d}$$

$$E_A\left(\boldsymbol{p}_S, \boldsymbol{p}_J, \boldsymbol{Q}, \boldsymbol{g}\right) \leqslant \gamma, \tag{14e}$$

where $P_S^{\max}$ and $P_J^{\max}$ are the maximal transmit power of the server and the jammer, respectively. Constraint (14a) limits the transmit power of the server at each time slot. (14b) indicates that the jamming signal power follows an uniformed distribution. Constraint (14c) limits the number of triples that can be transmitted per time slot and each triple can only be transmitted once. Since only partial time slots are used for semantic transmission, it is possible that some slots are not used, i.e., $\sum_{k=1}^{K} q_k^n = 0$. (14d) implies that the attacker can detect semantic information at most G time slots. Constraint (14e) is the semantic communication privacy requirement. From (14), we can see that the proposed optimization problem is non-convex and the objective function and constraint (14e) depends on scene graph embedding neural network model. Hence, the

relationship between the objective function and optimization variables cannot be represented exactly. In consequence, the problem (14) cannot be solved by traditional optimization algorithms.

III. Prioritized Sampling Assisted Twin Delayed Deep Deterministic Policy Gradient Algorithm

We introduce a prioritized sampling assisted twin delayed deep deterministic policy gradient algorithm (PS-TD3) to solve the problem in (14). Compared to current RL methods [37], the proposed algorithm can accurately estimate Q value function by clipped double Q learning so as to stably improve semantic quality of the received semantic information of the user and system privacy. Moreover, the introduced prioritized sampling technique can enable the agent to learn more from surprising and unexpected historical experience such that the training speed of the proposed method is improved. Next, we will first introduce the components of the proposed algorithm, and then explain the training process of the algorithm.

A. Components of PS-TD3 algorithm

In this section, we introduce the fundamental components of the proposed PS-TD3 algorithm as follows:

- *Agent*: The agent is the server that consecutively decides the triple to transmit, and power that is used for this transmission, at each time slot.
- *States*: The state captures the mutual importance distribution of all semantic triples, such that the state at time slot n is defined as a vector $\boldsymbol{s}_n = [\boldsymbol{\xi}_1^n, \ldots, \boldsymbol{\xi}_k^n, \ldots, \boldsymbol{\xi}_K^n]$, with $\boldsymbol{\xi}_k^n = \left[\cos\left(C\left(\boldsymbol{\psi}^k\right), C\left(\boldsymbol{\psi}^1\right)\right), \ldots, \cos\left(C\left(\boldsymbol{\psi}^k\right), C\left(\boldsymbol{\psi}^K\right)\right)\right]$ being the importance distribution of semantic triple $\boldsymbol{\psi}^k$ at time slot n. The sequence of server states recorded until the completion of semantic triple transmission is captured by a vector $[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \ldots, \boldsymbol{s}_N]$. Based on such mutual importance distribution, the server can find higher important triples that with higher cosine similarity for the most of triples. Hence, the server will put higher priority of this important triples in transmission so as to improve communication system privacy. Meanwhile, since the server does not know the transmit power of jamming signals, the server state does not include the transmit power of jamming signals.
- *Actions*: The action of server consists of choosing partial semantic triples to transmit and determining the corresponding transmit power at each time slot. In particular, at time slot n, the server action is $\boldsymbol{a}_n = (\boldsymbol{q}^n, p_S^n)$, with $\boldsymbol{q}^n$ being the transmission vector and p_S^n being the corresponding transmit power. The sequence of actions select over the whole semantic triple transmission process is $[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n, \ldots, \boldsymbol{a}_N]$
- *Reward*: The reward of the server capture the benefits of a selected action in terms of semantic transmission quality and system privacy. To achieve higher reward, the agent is required to maximize semantic similarity of the user and meanwhile minimize semantic similarity of the attacker, such that the reward at each step n is

$$r_n = \left(E_U^n - E_U^{n-1} - E_A^n + E_A^{n-1}\right) \mathbb{1}_{\left\{E_A^n \leqslant \gamma\right\}}, \quad (15)$$

where E_U^n is the temporal GNT of the received semantic information at the user, E_A^n is the temporal GNT of the eavesdropped semantic information, at the n-th time slot. $\mathbb{1}_{\left\{E_A^n \leqslant \gamma\right\}}$ is the indicator of the system privacy level, with $\mathbb{1}_{\left\{E_A^n \leqslant \gamma\right\}} = 1$ implying that the semantic transmission is secure at time slot n, and $\mathbb{1}_{\left\{E_A^n \leqslant \gamma\right\}} = 0$, otherwise. From (15), we see that the temporal reward will be zero with under-threshold privacy performance.

The sequence of state-action-reward transition captured until the completion of semantic triple transmission can be defined in a vector $[\boldsymbol{s}_1, \boldsymbol{a}_1, r_1, \boldsymbol{s}_2, \boldsymbol{a}_2, r_2, \ldots]$, which is referred as a trajectory. The cumulative reward of one complete semantic transmission trajectory is given as

$$\begin{aligned} R &= \begin{cases} \sum_{n=1}^{N} r_n + \gamma, & \text{if } \mathbb{1}_{\left\{E_A^N \leqslant \gamma\right\}}, \\ -\eta, & \text{otherwise}, \end{cases} \\ &= \begin{cases} E_U^N - E_A^N + \gamma, & \text{if } \mathbb{1}_{\left\{E_A^N \leqslant \gamma\right\}}, \\ -\eta, & \text{otherwise}, \end{cases} \end{aligned} \quad (16)$$

where $\eta > 0$ is a constant penalty factor. From (16), we can see that when the transmission is not private, the cumulative reward will be negative, i.e., $-\eta$. Otherwise (i.e., $\mathbb{1}_{\left\{E_A^N \leqslant \gamma\right\}}$), the cumulative reward will be non-negative, i.e. $\gamma - E_A^N \geqslant 0$. In other words, this cumulative reward will drive the agent to avoid semantic triple leakage for additional reward.

- *Deterministic Policy*: The deterministic policy is a mapping from a given state $\boldsymbol{s}_n$ to a deterministic action $\boldsymbol{a}_n$. The policy is implemented by the DNN based function approximator, i.e., actor network parameterized by $\boldsymbol{\phi}$, which establishes the relation between the mutual importance distribution of all semantic triples, the GNT of the user, and the semantic privacy. Compared to the stochastic policy that is the conditional probability of the agent choosing an action in a given state, the deterministic policy can converge with less samples by efficiently policy gradient estimation [38]. Specifically, the deterministic policy of the agent taking action $\boldsymbol{a}_n$ in a given state $\boldsymbol{s}_n$ can be expressed as $\boldsymbol{\pi}_{\boldsymbol{\phi}}(\boldsymbol{s}_n) = \boldsymbol{a}_n$.
- *Q Value function*: The Q value function $Q_{\boldsymbol{\theta}}(\boldsymbol{s}_n, \boldsymbol{a}_n)$ of the server is approximated by a DNN parameterized by $\boldsymbol{\theta}$, i.e. a critic network, which is used to estimate the expected future reward at state $\boldsymbol{s}_n$, given action $\boldsymbol{a}_n$. The input of this critic network is a state-action pair and the output is expected future reward.

  However, as studied in [39], the output of such critic network can be overestimated. In particular, the local-optimal actions yielding high estimated Q value will be assigned with high probability during action exploration within traditional deep Q learning based RL algorithms. Such local-optimal actions will dominate the estimation of expected future reward, i.e., Q value, and will stay overestimated, which disturbs the action exploration and

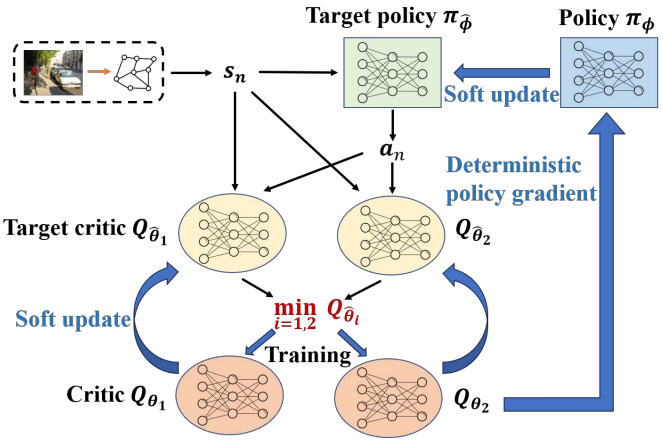


Fig. 4. The training process of the proposed PS-TD3 algorithm.

keeps the agent policy local-optimal. To solve such challenge, we merge the concept of clipped double Q learning, delayed policy update, and prioritized sampling into the training process of the PS-TD3 algorithm.

B. Training of PS-TD3 algorithm

Next, we will start the introduction of PS-TD3 training process with detailed explanation of the general training process of actor critic RL algorithm. The implementation of clipped double Q learning, the delayed policy update, and the prioritized sampling within this training process will be explained later.

Generally, starting from randomly initialized policy and critic networks, the training of RL algorithms includes two stages:

1) Interaction stage for experience collection: The server feeds its current environmental observation (i.e. states) into the policy network, which outputs deterministic action choices on the transmission time slot selection and transmit power control for each triple. Then, the reward is calculated based on the received and the eavesdropped semantic information of the user and the attacker. After a number of interactions between the agent and environment, a set of transitions that consist of state, action, reward, and next state, i.e., $(\boldsymbol{s}_n, \boldsymbol{a}_n, r_n, \boldsymbol{s}_{n+1})$, will be collected as historical experience and stored into a replay buffer $\mathcal{T}$.

2) Experience replay stage for networks update: In this stage, the agent will randomly sample a series of transitions from the replay buffer to update actor and critic networks. Specifically, the critic network $Q_{\boldsymbol{\theta}}$ will be updated based on the temporal difference (TD) based loss function

$$J(\boldsymbol{\theta}) = \frac{1}{N}\sum_{n=1}^{N}\left[y_n - Q_{\boldsymbol{\theta}}\left(\boldsymbol{s}_n, \boldsymbol{a}_n\right)\right]^2, \tag{17}$$

where n is the index of transition, $y_n = r_n + Q_{\hat{\boldsymbol{\theta}}}\left(\boldsymbol{s}_{n+1}, \boldsymbol{a}_{n+1}\right)$ is the learning target with $\boldsymbol{s}_{n+1}$ being the next state and $\boldsymbol{a}_{n+1}$ being next action. Here, the estimated Q value of the next state-action pair is calculated by the target Q function $Q_{\hat{\boldsymbol{\theta}}}(\cdot)$ that is softly updated by $\hat{\boldsymbol{\theta}} \leftarrow \tau\boldsymbol{\theta} + (1-\tau)\,\hat{\boldsymbol{\theta}}$ with τ being temperature parameter that adjusts updating rate. Then, based on updated critic network, the policy network can be updated by the deterministic policy gradient as in

$$\nabla_{\boldsymbol{\phi}} J(\boldsymbol{\phi}) = \frac{1}{N}\sum_{n=1}^{N} \nabla_{\boldsymbol{a}_n} Q_{\boldsymbol{\theta}}\left(\boldsymbol{s}_n, \boldsymbol{a}_n\right)|_{\boldsymbol{a}_n = \pi_{\boldsymbol{\phi}}(\boldsymbol{s}_n)} \nabla_{\boldsymbol{\phi}} \pi_{\boldsymbol{\phi}}\left(\boldsymbol{s}_n\right), \tag{18}$$

where $\boldsymbol{\phi}$ is the vector of policy network parameters.

This two stages will be implemented at the server alternately until a convergence is reached. The server will keep using the converged policy to make decisions on the transmission time slot and transmit power management for the covert semantic communication. As we mentioned in the former section, traditional actor-critic RL algorithms can suffer from overestimation of Q value. In what follows, we will treat this problem with the clipped double Q learning technique, as shown in Fig. 4, and will also merge the concept of delayed policy update and prioritized sampling in this two stage training process for stabilized and high time efficient training performance.

- **Relieved overestimation with clipped double Q learning.** With clipped double Q learning, we construct two critic networks $Q_{\boldsymbol{\theta}_1}$ and $Q_{\boldsymbol{\theta}_2}$, and only the smaller output Q value of these two target critic networks at each step will be chosen for TD calculation in (17). In other words, the learning target y_n is reconstructed as

$$y_n = r_n + \min_{i\in\{1,2\}} Q_{\hat{\boldsymbol{\theta}}_i}\left(\boldsymbol{s}_{n+1}, \boldsymbol{a}_{n+1}\right), \tag{19}$$

  where $\hat{\boldsymbol{\theta}}_i$ is the parameter of the target critic network $i \in \{1,2\}$. In this way, the experience replay stage is revised to be more conservative on taking dominating actions into the update of critic networks, so as to avoid overestimation.
- **Smoothed training process with delayed policy update.** Since the training of critic and policy networks are interdependent in the two training stages, we can expect that the training of policy can be unstable with unstable Q value output by target critic networks. Hence, to train the policy with a Q value estimated with lower variance, we propose to I) lower the update frequency of the policy network compared to critic networks. After a fixed number of updates to the critic networks, the policy network will be updated; II) add a clipped Gaussian noise μ to each output action, so as to enhance the exploration of the semantic communication environment with more diverse interaction and smoother estimation of Q value distribution; and III) introduce an additional target policy network $\pi_{\hat{\boldsymbol{\phi}}}$ whose parameters $\hat{\boldsymbol{\phi}}$ are soft updated by $\hat{\boldsymbol{\phi}} \leftarrow \tau\boldsymbol{\phi} + (1-\tau)\,\hat{\boldsymbol{\phi}}$ for further relieved overestimation.
- **Boosted training speed with prioritized sampling.** Since both of the used pessimistic Q value estimation and the delayed policy update is conservative learning strategy that may need more iteration to converge compared to traditional RL, we need to further expedite the training process. In particular, we proposed to integrate a prioritized sampling method into the interaction stage

of the training process to make fully use of the collected transitions. For standard interaction, the agent randomly samples a set of historical experience, i.e., transitions, from the replay buffer for RL training, considering the importance of each transition for training RL to be the same. However, in practice the contribution of different transitions can be different. For example, the transition with higher TD error, i.e., $\delta_n = y_n - Q_{\boldsymbol{\theta}}(\boldsymbol{s}_n, \boldsymbol{a}_n)$, will lead to a larger updating step. This is because the higher TD error indicates that the corresponding transition is more surprising and unexpected. Hence, the agent can learn more about Q value estimation from these transitions [40]. Therefore, we introduce a TD error related priority to improve the probability of sampling the transitions with high TD error. In particular, the probability of sampling transition with index n can be given by

$$\mathbf{Pr}(m) = \frac{b_n}{\sum_i b_i} = \frac{|\delta_n|^\alpha}{\sum_i |\delta_i|^\alpha} \tag{20}$$

where $b_n = |\delta_n|^\alpha$ is the priority factor of the transition, α is a constant to adjust priority. The impacts from the prioritized sampling will be trivial when we set a small α. Especially, $\alpha = 0$ represents that no priority is considered in sampling. By integrating (20) into the interaction stage, the agent is more likely to sample the transition with higher TD error, which speedup the convergence of the PS-TD3 algorithm. The specific training procedure including the clipped double Q learning and the prioritized sampling is summarized in **Algorithm 1**.

Algorithm 1 PS-TD3 algorithm for solving problem (14).

Initialize: Actor and critic networks parameters $\boldsymbol{\phi}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, Target networks parameters $\hat{\boldsymbol{\phi}} = \boldsymbol{\phi}, \hat{\boldsymbol{\theta}}_1 = \boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2 = \boldsymbol{\theta}_2$, Target network update rate τ, Replay buffer $\mathcal{T}$, Priority adjustment parameter α.
for $u = 1 \rightarrow U$ **do**
  for each environment step **do**
    Choose action $\boldsymbol{a}_n = \pi_{\hat{\boldsymbol{\phi}}}(\boldsymbol{s}_n) + \mu$ based on current target actor network.
    Calculate the step reward $r_n(\boldsymbol{s}_n, \boldsymbol{a}_n)$.
    Calculate TD error $\delta_n = r_n + \min_{i\in\{1,2\}} Q_{\hat{\boldsymbol{\theta}}_i}(\boldsymbol{s}_{n+1}, \boldsymbol{a}_{n+1}) - Q_{\boldsymbol{\theta}}(\boldsymbol{s}_n, \boldsymbol{a}_n)$ and priority factor $b_n = |\delta_n|^\alpha$.
    Collect transition into the buffer $\mathcal{T} = \mathcal{T} \cup \{(\boldsymbol{s}_n, \boldsymbol{a}_n, r_n(\boldsymbol{s}_n, \boldsymbol{a}_n), \boldsymbol{s}_{n+1}, b_n)\}$.
  end for
  for each network update step **do**
    Sample a batch of N transitions by the priority b_n.
    Update critic networks $Q_{\boldsymbol{\theta}_1}, Q_{\boldsymbol{\theta}_2}$ by $\nabla_{\boldsymbol{\theta}_i} J(\boldsymbol{\theta}_i)$.
    if a fixed number updating of critics have been done **then**
      Update actor network $\boldsymbol{\phi}$ by $\nabla_{\boldsymbol{\phi}} J(\boldsymbol{\phi})$.
      Update target actor network by $\hat{\boldsymbol{\phi}} \leftarrow \tau\boldsymbol{\phi} + (1-\tau)\hat{\boldsymbol{\phi}}$.
      Update target critic networks by $\hat{\boldsymbol{\theta}}_i \leftarrow \tau\boldsymbol{\theta}_i + (1-\tau)\hat{\boldsymbol{\theta}}_i$.
      Initialize the updating number of critic networks to 0.
    end if
  end for
end for

C. Complexity of the Proposed Algorithm

In this section, we analyze the complexity of the proposed PS-TD3 algorithm for covert semantic communication. The complexity of the PS-TD3 algorithm in training stage lies in actor and critic networks updating. In particular, the time-complexity of training each critic network that depends on the the number of hidden layers and neurons in each layer is given by

$$\mathcal{O}\left(U_c N \sum_{j=1}^{L_c-1} \omega_j \omega_{j+1} + U_c N\left(K^2+2\right)\omega_1 + U_c N \omega_{L_c}\right), \tag{21}$$

where U_c is the number of updating critic network, N is the number of transitions in one batch, w_j is the number of neurons in the i-th hidden layer, L_c is the number of hidden layers in critic network, K^2+2 is the input dimension of critic network that equals to the sum of dimensions of state $\boldsymbol{s}_n$ and action $\boldsymbol{a}_n$. Since the critic network is used to evaluate a deterministic policy, the output is a Q value scalar, i.e., the output dimension is one. Hence, the computation complexity at last output layer is ω_{L_c}. Similarly, we can obtain the time-complexity of training actor network, which is given by

$$\mathcal{O}\left(U_v N \sum_{j=1}^{L_v-1} \omega_j \omega_{j+1} + U_v N K^2 \omega_1 + 2U_v N \omega_{L_v}\right), \tag{22}$$

where U_v is the number of updating actor network, L_v is the number of hidden layers in actor network. Compared to the critic network, the input dimension and output dimension of actor networks equal to the dimension of state and action, respectively. Furthermore, from the introduction of training the proposed algorithm, we can see that the training procedure of target networks is partial parameter coping. Hence, we can compute the time-complexity of updating target actor and critic networks are $\mathcal{O}\left(U_v \sum_{j=1}^{L_v} \omega_j + U_v K^2 + 2U_v\right)$ and $\mathcal{O}\left(U_v \sum_{j=1}^{L_c} \omega_j + U_v K^2 + 2U_v\right)$, respectively. The proposed algorithm is trained offline. Therefore, after training, we only need to manage transmission time slot and power for private semantic communication by using actor network.

IV. Simulation Results and Analysis

In our simulations, we consider a circular wireless network where one server that locates at the center transmits image data to a randomly distributed user using semantic communication technique. At the same time, a randomly distributed attacker aims to eavesdrop the semantic information while a friendly jammer is deployed to protect the semantic transmission. The jammer constantly transmits jamming signal with random power. Here, we consider the attacker eavesdrops the transmission based on two kinds of power detectors in [34]

TABLE II: System Parameters

Parameter	Value	Parameter	Value
N	12	W	2 KHz
β	2	T	200 ms
σ_U^2	-30 dBm	σ_A^2	-30 dBm
$p_S^{\max}$	1 W	$p_J^{\max}$	1 W
G	8	γ	0.5
η	1	α	2
B	1		

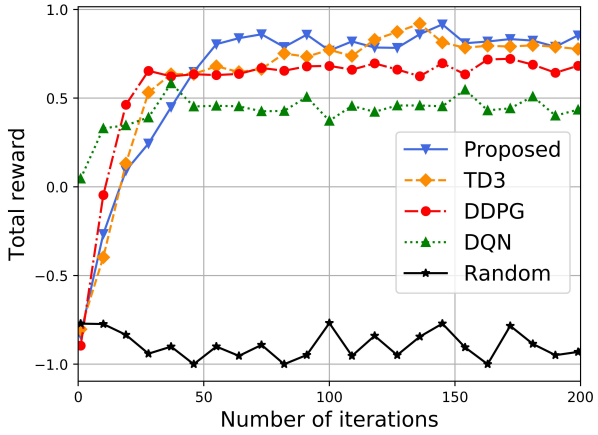


Fig. 5. Convergence of the proposed method.

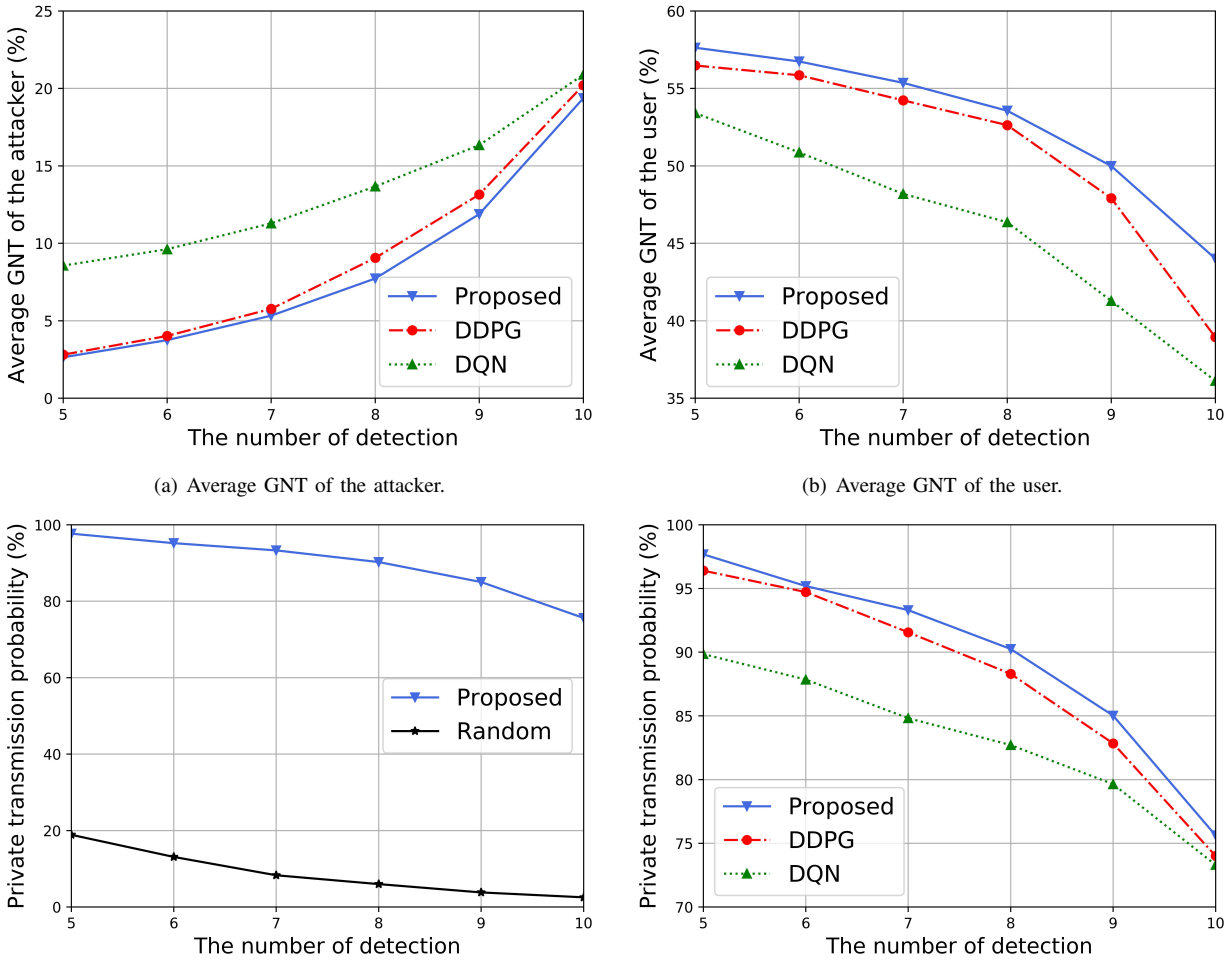
and [35]. In particular, the detector in [34] is more accurate, which requires to know current ratio of transmit power and jamming signal power. The detector in [35] only requires to know the distribution of the jamming signal power. The detection number of the attacker is G. Other system parameters are listed in Table II. We use the scene graph generation model in [33] for semantic information extraction and the scene graph embedding module in [41] for vectorization of semantic information. For comparison purposes, we consider four baselines:

- *Deep Q learning network (DQN) algorithm*. The comparison between the proposed algorithm and DQN is to justify how the clipped Q function of the proposed method can improve the semantic communication privacy performance.
- *Deep deterministic policy gradient (DDPG) algorithm*. The comparison between the proposed algorithm and the DDPG algorithm is to justify how the double Q function and the delayed policy update of the proposed method can address the problem of overestimating Q values.
- *Twin delayed deep deterministic policy gradient (TD3) algorithm*. The comparison between the proposed algorithm and the TD3 algorithm is to justify how the introduced prioritized sampling of the proposed method can reduce the training iterations.
- *Random method that randomly chooses time slot and transmit power for semantic transmission*. The comparison between the proposed algorithm and the random method is to justify how the proposed RL method enables the server to effectively optimize semantic transmission so as to improve system privacy level.

Figure 5 shows how the total reward of all considered solutions changes as the number of learning iterations varies. In Fig. 5, we can observe that compared with DQN and DDPG methods, the proposed algorithm can achieve about 77.8% and 14.3% improvement in terms of the reward, respectively, which can be directly reflected on higher system privacy and semantic quality of the received semantic information of the user. Such improvement stems from the fact that the use of clipped double Q function and delayed policy within the proposed algorithm encourages the agent to escape from local optimal actions and keep searching more advantageous ones. From the Fig. 5, we can also see that compared to the plain TD3 algorithm that requires about 80 iterations to converge, the proposed PS-TD3 algorithm reaches convergence after approximately 50 iterations. This is because the prioritized sampling mechanism introduced in the proposed method enables the agent to focus on more impactful transitions, thereby accelerating the training process.

Figure 6 shows how the joint communication and privacy performance of the proposed algorithm varies with the attacker's detection number. Overall, from Fig. 6, we can see that as the detection number increases, the quality of semantic information eavesdropped by the attacker is improved such that the system's privacy crisis intensifies. This is because the server will tend to transmit more conservatively to avoid privacy leakage, e.g., lower transmit power, such that decreasing the transmission performance at the user. Meanwhile, Fig. 6a) shows that the proposed method keeps its average semantic similarity of eavesdropped semantic information lower than the other ones of baseline methods with the increasing detection numbers. Then, from Fig. 6b), we can see that the proposed method can acquire higher GNT of the user as the number of detection varies. This stems from the fact that the proposed algorithm enables the agent to aggressively search for the global optimal transmit power that can hide existence of transmission from frequent eavesdropping attacks. Finally, Fig. 6c) and Fig. 6d) shows how the probability of private semantic transmission achieved by all considered methods changes with the attacker's detection number. From Fig. 6c), we can see that the proposed method can significantly improve the privacy of = semantic communication system compared to the random methods, which justify the effectiveness of the proposed RL based power control scheme. Meanwhile, from Fig. 6d), we observe that the proposed algorithm can achieve higher level of system privacy, compared to traditional RL methods. These gains stem from the fact that the combination of the clipped double Q function and delayed policy update in the proposed method enables the agent to aggressively search for the most effective transmission policy steadily without being mislead from Q value overestimation.

Figure 7 shows how joint communication and privacy performance of the proposed PS-TD3 algorithm varies with the number of transmission time slots. In general, Fig. 7 demonstrates that the semantic transmission become more private as the number of available transmission time slots increases, and the quality of the semantic information received



(a) Average GNT of the attacker.

(b) Average GNT of the user.

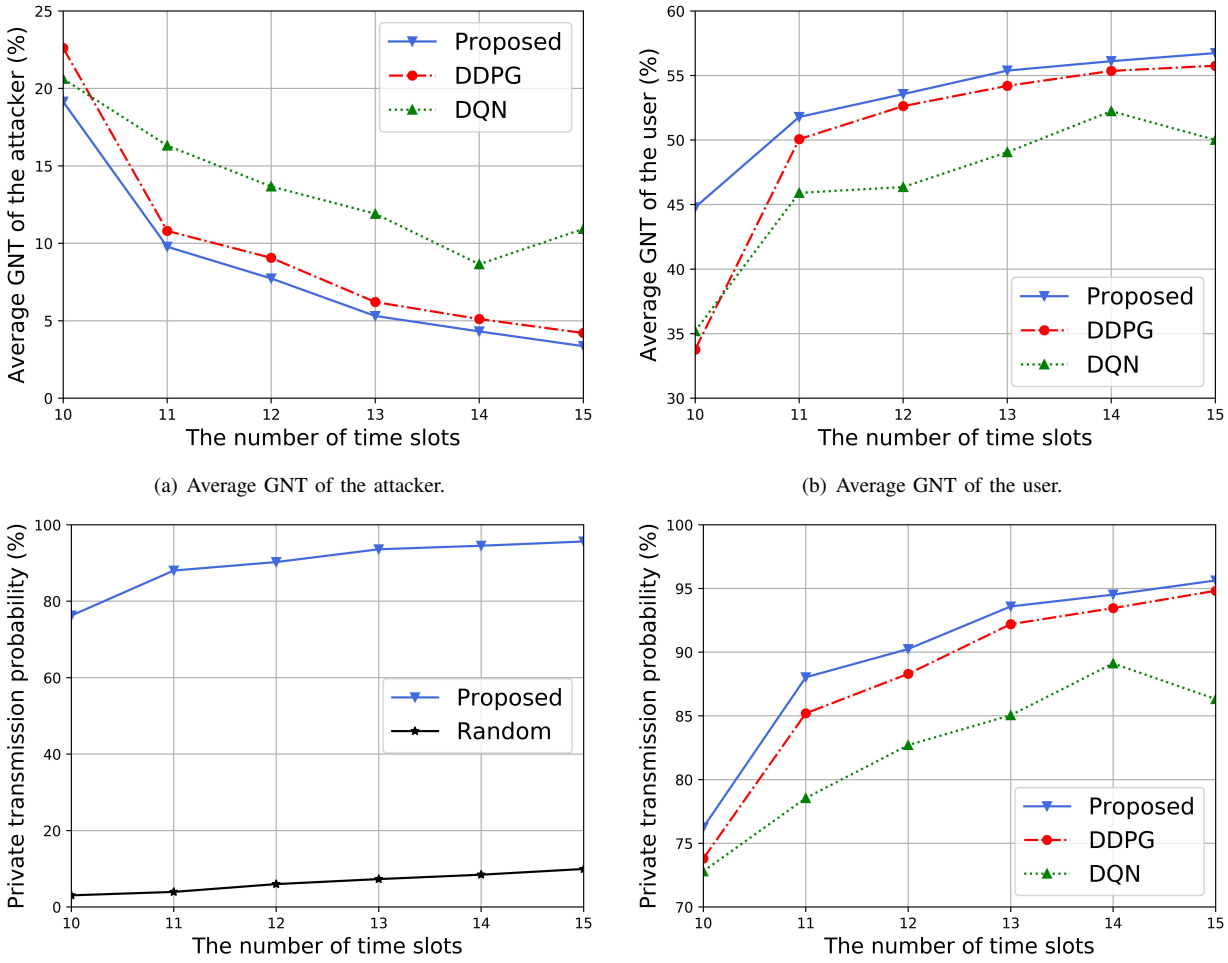
(c) Private transmission probability versus random method.

(d) Private transmission probability versus traditional RL methods.

Fig. 6. Private semantic transmission performance as detection number varies.

by the user is also improved. This is because there will be more flexible choices for the server as the available time slots increasing. In Fig. 7a), we can observe that the proposed PS-TD3 algorithm can keep the quality of the eavesdropped semantic information in a low level, e.g., semantic similarity lower than 20%. At the same time, from Fig. 7b), we can see that the proposed method can also acquire higher GNT of the received semantic information of the user as the number of transmission time slots changes. In Figs. 7c) and 7d), we can also observe that the proposed method can significantly improve the system privacy compared to baseline methods. In particular, from Figs. 7a), 7b), and 7d), we can see that the DQN baseline method suffers non-trivial performance loss as the available transmission slots increases, i.e., enlarged action space. However, the proposed PS-TD3 algorithm can stably improve the semantic communication performance with the additional time slots. This is because the proposed PS-TD3 algorithm uses clipped double Q function to keep the server aggressively searching for global optimal solutions within the large action space, and uses the delayed updated deterministic policy to stabilize such searching. Finally, Figs. 7a), 7b), and 7d) also show a great performance loss at all considered RL baseline methods (i.e. DQN and DDPG) with reduced number of available time slots for transmission, while the proposed PS-TD3 algorithm can consistently achieve better performance in semantic quality and privacy as the number of time slots changes. The reason is that the accurately estimated Q value of the proposed RL method enable the agent to learn truly valuable actions.

Figure 8 shows how the transmission performance varies as the number of available transmission slots increases, given that the attacker uses the power detector in [35]. Here, we note that, in this scenario, the privacy can be guaranteed by all the considered RL methods (private transmission probability over 90%). From Fig. 8a), we can see that the average GNT of the user increases with the available transmission time slots. Specifically, compared to the other two RL baseline methods, the quality of the received semantic information of the proposed PS-TD3 algorithm is significantly higher. Figure 8b) shows how the percentage of transmitted semantic triples changes with the number of transmission slots. A higher percentage means that more semantic triples can be transmitted, such that higher quality of semantic information can be achieved. From Fig. 8b), we can see the transmitted semantic triples percentage of the proposed algorithm increases faster than DDPG method and DQN method. This



(a) Average GNT of the attacker.

(b) Average GNT of the user.

(c) Private transmission probability versus random method.

(d) Private transmission probability versus traditional RL methods.

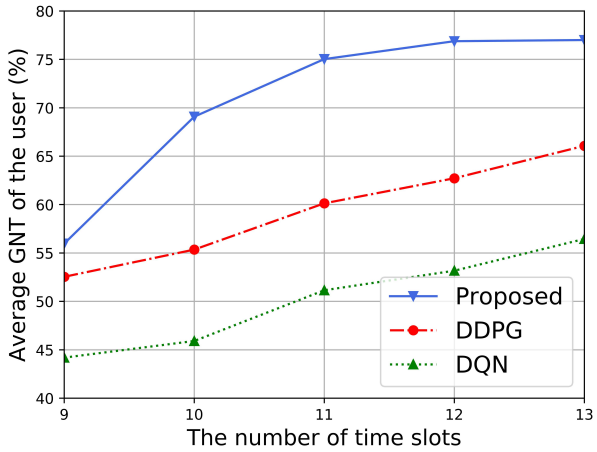
Fig. 7. Private semantic transmission performance as the number of transmission slots varies.

is due to fact that the proposed PS-TD3 algorithm enables the server to effectively take advantage of additional transmission time slots to support more semantic triples transmission when confronts to a less powerful attacker, so as to improve semantic information quality of the user.

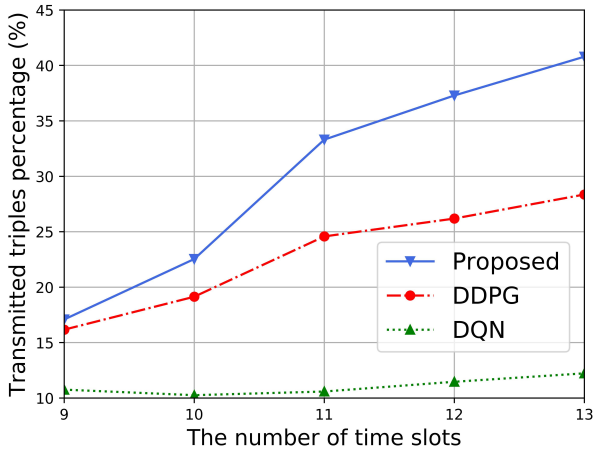
Figure 9 shows the partial semantic triples transmission management at each time slot of the proposed method and baseline RL methods. The color of blocks represent the corresponding frequencies that computed based on 10,000 transmissions. In particular, as the frequency of detection at each time slot increases, the color of the first row in figs. 9a), 9b), and 9c) changes from white to green. Similarly, the color of the second row in fig. 9 changes from white to green as the frequency of transmission at each time slot decreases (i.e., the frequency of no transmission increases). Then, the color changes of blocks in the third rows at fig. 9 represent the absolute value of difference between the blocks in the first row and the second row, i.e., the white block is corresponding to better privacy to avoid detection while the green block represents worse transmission protection. From figs. 9a), b), and c), compared to baseline RL methods, we can see that the proposed PS-TD3 algorithm can enable the server to properly arrange partial semantic triples transmissions at each time slot to avoid detection.

V. Conclusion

In this paper, we have developed a novel covert semantic communication framework that consistently secures the semantic image transmission between a server and a user from eavesdropping attacks by jointly optimizing the semantic triple selection and transmit power control of each semantic transmission of the server. Within this framework, an independent friendly jammer is deployed to protect the semantic transmission without inter-device communication and cooperation for reduced spectrum abuse, energy costing, and information leakage. We have proposed a GNT metric to evaluate the communication performance and system privacy, and have cast this semantic triple selection and transmit power control problem in an optimization setting. We have also introduced a prioritised sampling assisted, twin delayed deep deterministic policy gradient RL solution to solve this non-convex problem in low computational and space complexity. Simulation results have shown that the proposed algorithm achieves high system privacy level and better quality of the received semantic information with fast convergence.

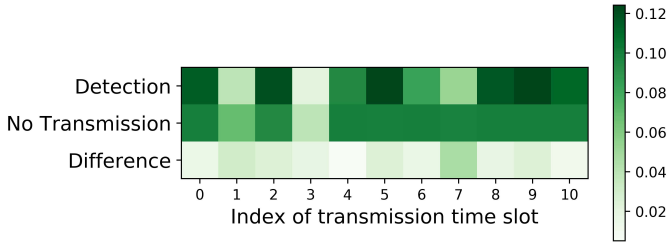


(a) Average GNT of the user.

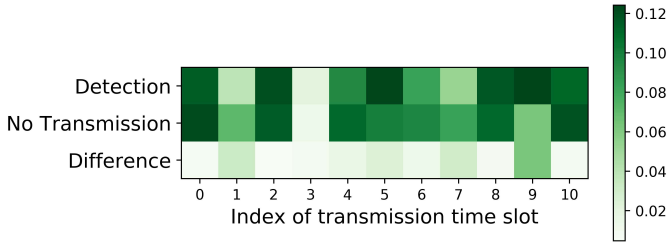


(b) Transmitted semantic triples percentage

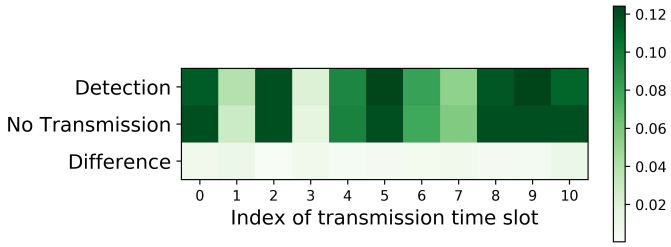
Fig. 8. Transmission performance confronts the power detector in [35].



(a) Transmission slots arrangement of DQN method.



(b) Transmission slots arrangement of DDPG method.



(c) Transmission slots arrangement of the proposed method.

Fig. 9. Transmission time slot management.

REFERENCES

[1] M. Hua, Q. Wu, W. Chen, O. A. Dobre, and A. L. Swindlehurst, "Secure intelligent reflecting surface-aided integrated sensing and communication," *IEEE Transactions on Wireless Communications*, vol. 23, no. 1, pp. 575–591, Jun. 2024.

[2] Y. Hu, M. Chen, and W. Saad, "Joint access and backhaul resource management in satellite-drone networks: A competitive market approach," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 3908–3923, Mar. 2020.

[3] W. Yang, H. Du, Z. Liew, W. Lim, Z. Xiong, D. Niyato, X. Chi, X. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213–250, Nov. 2023.

[4] D. Deng, C. Wang, L. Xu, F. Jiang, K. Guo, Z. Zhang, W. Wang, T. Q. S. Quek, and P. Zhang, "Semantic communication empowered NTN for IoT: Benefits and challenges," *IEEE Network*, vol. 38, no. 4, pp. 32–39, Apr. 2024.

[5] Y. Liu, H. Du, D. Niyato, J. Kang, Z. Xiong, S. Mao, P. Zhang, and X. Shen, "Cross-modal generative semantic communications for mobile AIGC: Joint semantic encoding and prompt engineering," *IEEE Transactions on Mobile Computing*, pp. 1–16, Aug. 2024.

[6] C. Xing, J. Lv, T. Luo, and Z. Zhang, "Representation and fusion based on knowledge graph in multi-modal semantic communication," *IEEE Wireless Communications Letters*, vol. 13, no. 5, pp. 1344–1348, Feb. 2024.

[7] J. Zhang, M. Chen, Y. Zhu, H. Shihao, T. Luo, and Z. Zhang, "Performance optimization of semantic communications for users with heterogeneous knowledge," in *Proc. IEEE International Conference on Communications*, Denver, CO, USA, Jun. 2024, pp. 5515–5520.

[8] J. Zhao, M. Chen, Z. Yang, C. You, and M. Chen, "Resource allocation for semantic relay aided wireless networks with probability graph," in *Proc. IEEE International Conference on Communications*, Denver, CO, USA, Jun. 2024, pp. 5317–5322.

[9] T. M. Getu, W. Saad, G. Kaddoum, and M. Bennis, "Performance limits of a deep learning-enabled text semantic communication under interference," *IEEE Transactions on Wireless Communications*, vol. 23, no. 8, pp. 10213–10228, Mar. 2024.

[10] J. Hu, F. Wang, W. Xu, H. Gao, and P. Zhang, "SemHARQ: Semantic-aware HARQ for multi-task semantic communications," Available Online: https://arxiv.org/abs/2404.08490, 2024.

[11] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *IEEE Communications Surveys & Tutorials*, pp. 1–1, Jun. 2024.

[12] D. Gunduz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, and C.-B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, Nov. 2023.

[13] Z. Qin, L. Liang, Z. Wang, S. Jin, X. Tao, W. Tong, and G. Y. Li, "AI empowered wireless communications: From bits to semantics," *Proceedings of the IEEE*, pp. 1–32, Aug. 2024.

[14] P. Zhang, W. Xu, Y. Liu, X. Qin, K. Niu, S. Cui, G. Shi, Z. Qin, X. Xu, F. Wang, Y. Meng, C. Dong, J. Dai, Q. Yang, Y. Sun, D. Gao, H. Gao, S. Han, and X. Song, "Intellicise wireless networks from semantic communications: A survey, research issues, and challenges," *IEEE Communications Surveys & Tutorials*, pp. 1–1, Aug. 2024.

[15] H. Wu, Y. Shao, E. Ozfatura, K. Mikolajczyk, and D. Gunduz, "Transformer-aided wireless image transmission with channel feedback," *IEEE Transactions on Wireless Communications*, pp. 1–1, Apr. 2024.

[16] X. Peng, Z. Qin, X. Tao, J. Lu, and L. Hanzo, "A robust semantic text communication system," *IEEE Transactions on Wireless Communications*, pp. 1–1, Apr. 2024.

[17] Z. Yang, M. Chen, G. Li, Y. Yang, and Z. Zhang, "Secure semantic communications: Fundamentals and challenges," *IEEE Network*, pp. 1–1, Jun. 2024.

[18] T. Zuo, Y. Duan, Q. Du, and X. Tao, "Semantic security: A digital watermark method for image semantic preservation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2024, pp. 4645–4649.

[19] Y. Lin, H. Du, D. Niyato, J. Nie, J. Zhang, Y. Cheng, and Z. Yang, "Blockchain-aided secure semantic communication for AI-generated content in metaverse," *IEEE Open Journal of the Computer Society*, vol. 4, pp. 72–83, Mar. 2023.

[20] G. Nan, Z. Li, J. Zhai, Q. Cui, G. Chen, X. Du, X. Zhang, X. Tao, Z. Han, and T. Q. S. Quek, "Physical-layer adversarial robustness for deep learning-based semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2592–2608, Jun. 2023.

[21] X. Ren, J. Wu, H. Xu, and X. Chen, "Diffusion model based secure semantic communications with adversarial purification," in *IEEE 10th Conference on Big Data Security on Cloud*, NYC, NY, USA, Jun. 2024, pp. 130–134.

[22] Q. Qin, Y. Rong, G. Nan, S. Wu, X. Zhang, Q. Cui, and X. Tao, "Securing semantic communications with physical-layer semantic encryption and obfuscation," in *Proc. IEEE International Conference on Communications*, Rome, Italy, Oct. 2023, pp. 5608–5613.

[23] X. Liu, G. Nan, Q. Cui, Z. Li, P. Liu, Z. Xing, H. Mu, X. Tao, and T. Q. S. Quek, "SemProtector: A unified framework for semantic protection in deep learning-based semantic communication systems," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 56–62, Nov. 2023.

[24] Y. Chen, Q. Yang, Z. Shi, and J. Chen, "The model inversion eavesdropping attack in semantic communication systems," in *Proc. IEEE Global Communications Conference*, Kuala Lumpur, Malaysia, Dec. 2023, pp. 5171–5177.

[25] M. Zhang, Y. Li, Z. Zhang, G. Zhu, and C. Zhong, "Wireless image transmission with semantic and security awareness," *IEEE Wireless Communications Letters*, vol. 12, no. 8, pp. 1389–1393, May. 2023.

[26] X. Chen, J. An, Z. Xiong, C. Xing, N. Zhao, F. R. Yu, and A. Nallanathan, "Covert communications: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1173–1198, Apr. 2023.

[27] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, Aug. 2013.

[28] T. V. Sobers, B. A. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Transactions on Wireless Communications*, vol. 16, no. 9, pp. 6193–6206, Jun. 2017.

[29] R. Xu, G. Li, Z. Yang, M. Chen, Y. Liu, and J. Li, "Covert and reliable semantic communication against cross-layer privacy inference over wireless edge networks," in *Proc. IEEE Wireless Communications and Networking Conference*, Apr. 2024, pp. 1–6.

[30] J. Hu, L. Ye, Y. Chen, X. Zhang, J. Wang, and Z. Chen, "Covert communications for text semantic with finite blocklength," *IEEE Wireless Communications Letters*, pp. 1–1, Aug. 2024.

[31] Y. Wang, Y. Hu, H. Du, T. Luo, and D. Niyato, "Multi-agent reinforcement learning for covert semantic communications over wireless networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Rhodes Island, Greece, May. 2023, pp. 1–5.

[32] H. Du, G. Liu, D. Niyato, J. Zhang, J. Kang, Z. Xiong, B. Ai, and D. I. Kim, "Generative Al-aided joint training-free secure semantic communications via multi-modal prompts," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2024, pp. 12896–12900.

[33] K. Tang, Y. Niu, J. Huang, J. Shi, and H. Zhang, "Unbiased scene graph generation from biased training," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, Jun. 2020.

[34] M. Forouzesh, P. Azmi, N. Mokari, K. K. Wong, and D. Goeckel, "Information-theoretic security or covert communication," Available online: https://arxiv.org/abs/1803.06608, 2019.

[35] L. Yang, W. Yang, S. Xu, L. Tang, and Z. He, "Achieving covert wireless communications using a full-duplex multi-antenna receiver," in *Proc. IEEE 5th International Conference on Computer and Communications*, Dec. 2019, pp. 912–916.

[36] W. Zhang, Y. Wang, M. Chen, T. Luo, and D. Niyato, "Optimization of image transmission in cooperative semantic communication networks," *IEEE Transactions on Wireless Communications*, vol. 23, no. 2, pp. 861–873, Jun. 2024.

[37] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," Available online: https://arxiv.org/abs/1509.02971, 2019.

[38] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. International Conference on Machine Learning*. 2014, ICML'14, pp. 387–395, JMLR.org.

[39] S. Fujimoto, H. v. Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," Available Online: http://arxiv.org/abs/1802.09477, 2018.

[40] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," Available Online: https://arxiv.org/pdf/1511.05952, 2016.

[41] J. Johnson, A. Gupta, and F.-F. Li, "Image generation from scene graphs," Available online: http://arxiv.org/abs/1804.01622, 2018.