Deep Learning Enabled Semantic-Secure Communication with Shuffling

Fupei Chen*, Liyao Xiang*, Hei Victor Cheng†, and Kaiming Shen‡
*John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai, China
†Electrical and Computer Engineering Department, Aarhus University, Denmark
‡School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), China
E-mail: cfp2022@sjtu.edu.cn; xiangliyao08@sjtu.edu.cn; hvc@ece.au.dk; shenkaiming@cuhk.edu.cn

***Abstract*—Deep learning and natural language processing draw heavily on the recent progress in semantic communications; this paper examines the security aspect of this cutting-edge technique. Our goal is to improve upon the conventional secure coding methods to strike a superior tradeoff between transmission rate and leakage rate. Toward this end, we devise a novel semantic security communication system wherein the random shuffling pattern serves as the secret key shared. Intuitively, the permutation of words in the same text via shuffling would result in the meaning distortion of the target text to such a great extent that an eavesdropper can no longer recover the semantic truth. The proposed method can be rephrased as maximizing the transmission rate while minimizing the semantic error probability under the given leakage rate constraint. Simulations demonstrate the significant advantage of the proposed method over the benchmark in boosting secure transmission, especially when channels are prone to strong noise and unpredictable fading, can achieve up to 60% performance gain.**

***Index Terms*—Secure semantic communication, deep learning, shared key, wiretap channel.**

I. Introduction

Secure transmission over a wiretap channel has attracted considerable research interest since Shannon [1] initiated this subject in the 1940s. As a fundamental result in this area, it is shown that perfect secrecy can be realized for the wiretap channel so long as the signal received at the eavesdropper side is statistically independent of the information source, namely *information-theoretical security* [2]. The above result has been further extended in multiple respects, ranging from the capacity-reaching code design [3] to the secret key-sharing strategy [4].

In contrast to the fruitful studies on the theoretical aspect of secure transmission, the practical implementation and algorithm design are somewhat lagging behind. As a pioneering result, the one-time pad strategy due to Shannon [2] preserves the secrecy completely, but it is of limited practical value because the key string needs to be at least equally long as the information string. Ardestanizadeh et al. [5] aim to enhance Wyner' coding scheme by incorporating a fresh randomness feedback strategy, thus establishing the feedback capacity of the wiretap channel. The more recent works including [6], [7] suggest using Polar codes and LDPC for the secure transmission purpose, respectively. But they still incur high computational complexities.

Differing from the above traditional attempts, this work considers secure transmission from a semantic communication point of view. Semantic communication has worked its way into a new frontier in the realm of communication theory by sending the implicit meaning and intention behind the text rather than the text itself. This new technique is envisioned to significantly enhance transmission efficiency and reliability. In light of the tools of semantic communication, we devise a paradigm wherein the communication toward the legitimate receiver can be performed using much fewer bits than the traditional Shannon's scheme while satisfying the information leakage constraint. Intuitively speaking, in contrast to the conventional efforts that try to protect every single bit of the transmission string, this present work pursues secure protection only for those semantically essential bits; in other words, those bits recovered by an eavesdropper are insufficient to reveal any core content. Such a soft policy yields a substantial reduction of the secrecy-preserving overhead cost.

To the best of our knowledge, only a few existing works have considered security for semantic communications. [8] proposes a public key encryption scheme based on learning with error (LWE) generated key. [9] proposes a bilingual evaluation understudy (BLEU) fractional key generation mechanism and a subcarrier obfuscation mechanism based on OFDM to protect semantic data through encryption and obfuscation. A novel deep neural network (DNN) based framework tailored to semantic extraction forms the building block of this work. For text transmission, [10] proposes a bidirectional long short-term memory (BiLSTM) based coding scheme for semantic transmission, whereas [11] resorts to the Transformer for robust semantic transmission of text, and [12] concerns with the semantic transmission of speech, proposing DeepSC-S structured as nonlinear transform and conditional coding—which facilitate the extraction of semantic features of the speech data.

Our work stems from the crucial observation that DNNs can learn to use secret keys to protect information from other DNNs

Liyao Xiang (xiangliyao08@sjtu.edu.cn) is the corresponding author.

This work was partially supported by NSF China (62272306, 62032020, 62136006) and Shanghai Natural Science Fundation (21JC1403800), NSFC with Grant No. 62293482, the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone, and the Aarhus Universitets Forskningsfond project number AUFF 39001. Authors would like to appreciate the Student Innovation Center of SJTU for providing GPUs.

979-8-3503-1090-0/23/$31.00 © 2023 IEEE

[13] via end-to-end adversarial training. Following this line, we propose the use of machine translation technology (which is originated in the natural language processing area) for the physical layer security enhancement of semantic communication. Specifically, we devise a semantic joint source-channel coding scheme that extracts the semantic essence of raw information in the presence of random shuffling and channel noise.

The transmitter (resp. receiver) in our semantic communication system can be decomposed into the Transformer encoder (resp. decoder) and the multi-layer perceptron (MLP) encoder (resp. decoder). The intermediate string produced by the encoder is shuffled but then can be restored at the decoder. More specifically, the permutation pattern is somehow conveyed to the legitimate receiver as the shared secret key. The rationale is as follows: the semantic facts are sensitive to the positions of words. In other words, merely modifying the word positions can completely alter the meaning of the text. On the contrary, an eavesdropper who decodes the disarranged string without the key will face difficulty in recovering the original order—not to mention the semantic truth. With a mutual information and leakage rate based loss function, we train the above framework in an end-to-end fashion to learn the optimal coding that matches the given key. Furthermore, unlike adversarial training, the proposed shuffling scheme does not incur any additional overhead for learning.

As the main contribution of this work, we devise a semantically secure communication framework based on a novel idea of implementing the shared secret key by means of symbol shuffling. According to our simulations, the proposed framework is capable of enhancing the semantic communication while suppressing information leakage. In particular, much better robustness can be achieved in the low-SNR regime of the Rayleigh fading scenario by our semantic approach as compared to the traditional secure transmission methods.

II. System Model

We consider the point-to-point communication system with an eavesdropper shown in Fig. 1. The sender Alice wishes to communicate a message to the receiver Bob while keeping it secret from the eavesdropper Eve. The sender maps a sentence $S = [w_1, w_2, ...w_l]$, where $w_i, \forall i = 1, 2, ..., l$ is the ith word, into a symbol stream with a shared key K, and then passes it through the physical channel with transmission impairments such as distortion and noise. The receiver decodes the received with K to estimate the original sentence. Meanwhile, the eavesdropper also tries to decode the received to recover the sentence without K. The goal of the system is to minimize the semantic errors while reducing the number of symbols to be transmitted between legitimate sender and receiver, yet without leaking any information to an eavesdropper. Different from existing secure communication systems for bit-level transmission, we propose joint semantic and channel coding which permits transmission/leakage at the semantic level, i.e., if the semantic recovery fails due to bit loss, the transmission is considered unsuccessful.

To achieve successful semantic recovery and prevent leakage, we jointly design the sender and receiver with DNNs along with a shared key to enable legitimate transmission while suppressing eavesdropping. Particularly, each sentence is padded to a fixed length N to constitute the input $M = [w_1, w_2, ...w_N]$ which is further converted into $U \in \mathbb{R}^{N\times L}$ through an embedding layer. L suggests the dimension of the word vector. The input embedding is fed into a semantic encoder to output $T \in \mathbb{R}^{N\times L}$ which is sent to the channel encoder to produce $X \in \mathbb{R}^{N\times V}$ where V is the feature dimension. X is transmitted through the physical channel and is received as Y, a noisy version of X. The eavesdropper receives Z through a wiretap channel, and we do not assume it is a degraded channel. Symmetric to the sender, the receiver (the eavesdropper) feeds $Y(Z)$ through a channel decoder and then a semantic decoder to reconstruct the input $\hat{M}(M')$.

A. The Shared Key

Without using the shared key, both the receiver and the eavesdropper can successfully decode the received message. To prevent leakage, we design a key suitable to the DNN-based sender and receiver. The intuition is that the DNN intermediate representation is often of a high dimension, which is hard to recover if being randomly shuffled. Specifically, we randomly permute word vectors $U \in \mathbb{R}^{N\times L}$ by multiplying a permutation matrix $P_R \in \{0,1\}^{N\times N}$ such that $P_R U$ stands for a row-shuffled U. Correspondingly, column shuffling is $U P_C$ where $P_C \in \{0,1\}^{L\times L}$. Notably, the two shuffling methods have different indications that row permutation means rearranging word position within a sentence, whereas column permutation refers to the altering of the word.

Due to the permutation invariance of Transformers [14], we adopt Transformer encoder blocks and decoder blocks as the DNN structure for both the semantic encoder and decoder respectively, and MLP layers for the channel encoder and decoder. Considering the impact of shuffling, we restrict the position of column permutation to the output of the channel encoder, and row shuffling takes place at one of the following: 1) output of the embedding layer (U); 2) output of the semantic encoder (T); and 3) output of the channel encoder (X), as shown in Fig. 1. At the receiver, it only needs to multiply the corresponding inverse shuffling matrix at the symmetric position of the "encryption" to "decrypt" the received message. The three positions, as we will later illustrate, are equivalent to the receiver, but have minor impacts on the eavesdropper.

Random permutation matrices P_R and P_C both serve as the shared key between Alice and Bob. It can be perceived that, when the permutation order is uniformly chosen, all sequences are equally likely to be inverted to the original order. Hence without the shared key, Eve can hardly reconstruct the original message.

B. Secrecy Capacity and Key Rate

To describe the goal of secure communication between Alice and Bob, we use the secrecy capacity [15], referring to the

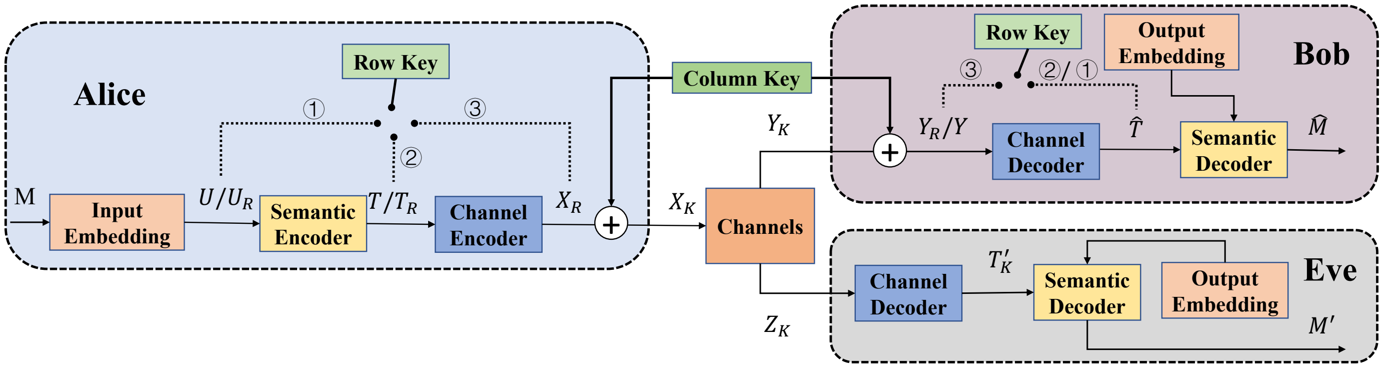


Fig. 1: Secure communication over a public channel with a shared key. The subscript R, K represents row shuffling, row and column shuffling are conducted, respectively. Input and output embedding refer to the same background knowledge.

maximum transmission rate achieved by legitimate parties with a shared key K, while ensuring near-zero information leakage rate R_L:

$$C_S(R_K) = \max_{p(x)} \left\{ I(X;Y) - [I(X;Z) - R_K]^+ \right\}, \tag{1a}$$

$$\text{s.t.} \lim \sup_{n\to\infty} R_L = \frac{1}{n} I(M;Z) \le \epsilon, \tag{1b}$$

where $I(\cdot;\cdot)$ is the mutual information, R_K denotes the transmission rate of the key, n is the length of symbol stream transmitted, and $[x]^+ = \max\{x, 0\}$. The achievability suggests that when the transmission rate $R < C_S$, there exists a coding scheme that allows the legitimate receiver to recover the original input with arbitrarily small error probability while preventing the information from being leaked to the eavesdropper.

Therefore, in information theoretic secrecy, the security of the communication system is quantified by the information leakage rate (R_L), while the secrecy capacity quantifies reliability.

In practice, it is hard to evaluate the mutual information in Eq. (1), and thus we resort to a neural estimator [16] to obtain an approximate evaluation. Specifically, mutual information can be rewritten as the Kullback-Leibler (KL) divergence between the joint probability density and the product of the marginal probabilities:

$$I(X;Y) = D_{\mathrm{KL}}(p(X,Y) \| p(X)p(Y)). \tag{2}$$

Based on the Donsker-Varadhan variational representation, the KL divergence has the following dual representation:

$$D_{KL}(\mathbb{P}\|\mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log\left(\mathbb{E}_{\mathbb{Q}}\left[e^T\right]\right) \tag{3}$$

where the supremum is taken over all functions T such that the two expectations are finite. For any given function $T : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$, the right-hand side of the aforementioned expression corresponds to a lower bound on mutual information $I(X;Y)$.

Similar to [11], a neural network T_θ parameterized by θ is used in substitution of T. Hence we maximize the lower bound to estimate mutual information:

$$L_{MI}^{\theta}(X;Y) = \mathbb{E}_{\mathbb{P}_{XY}}\left[T_\theta\right] - \log\left(\mathbb{E}_{\mathbb{P}_X \otimes \mathbb{P}_Y}\left[e^{T_\theta}\right]\right). \tag{4}$$

Aside from the mutual information, the key rate R_K in Eq. (1) also needs to be calculated. Note that the conventional unit for R_K is bits per symbol and the key length required for transmission is considered. In the realm of semantic secure communication, the basic unit of transmission is a word and shuffling order acts like the shared key. Thereby we have:

Proposition 1. *For the input embedding $U \in \mathbb{R}^{N\times L}$, the key rate is given by*

$$R_K = \frac{\log_2(N!L!)}{N}. \tag{5}$$

Proof: There are $N!$ possible row permutations and $L!$ possible column permutations with equal probabilities for the input. Therefore, the entire space for possible shared keys is $N! \cdot L!$ which could be described by a key of length $\log_2(N!L!)$. For each word, the average number of bits required to represent the key equals the length divided by N. ■

C. Loss Functions and Training

The communication system in Fig. 1 is trained in an end-to-end fashion. The semantic encoder learns to extract sentence meaning, compress it, and "encrypt" it by shuffling, whereas the semantic decoder learns to reconstruct the input by "decrypting" the symbol stream. Both channel encoder and decoder learn to adjust the data distribution to best match the channel condition. Without the key, Eve can only try to decipher the received message by trial and error. Let the network parameters for Alice, Bob, Eve be $\theta_A, \theta_B, \theta_E$, respectively. The training contains two parts: the legitimate channel between Alice and Bob, and the eavesdropping channel between Alice and Eve. By jointly training the encoding and decoding networks end to end, both θ_B and θ_E could influence θ_A through backpropagation.

The legitimate parties train over all input instances and randomly chosen permutation matrices. For example, if the row key is chosen at position 1 in Fig. 1, Alice shuffles the embedding U for M by a random permutation matrix P_R to obtain $U_R = P_R U$, and then feed U_R through the semantic and channel encoders to obtain X_R. It can be verified by the Transformer's property that $X_R = P_R X$ where X is the original output unshuffled. Ignoring the transmission impairments, the receiver could obtain X_R and feeds it into its channel decoder and semantic decoder to get M_R, which can be inverted by $\hat{M} = P_R^{-1} M_R$ with the row key. Notably, $\hat{M}$ is the reconstructed M without shuffling. It takes advantage of the permutation invariance of Transformers. In our case, we let Bob "decrypt" the received message in the same way despite the noisy channels. The column key is applied in a similar manner but only at both ends of the channel due to Transformer's property constraints. Putting shuffling in the loop, we reconstruct the input by minimizing the cross-entropy loss:

$$L_{CE} = CE(M; \hat{M}). \tag{6}$$

In addition to cross-entropy loss, $C_S(R_K)$ in Eq. (1) is included as a loss term to encourage raising transmission rate. In the meanwhile, the leakage rate R_L in Eq. (1) should be suppressed. The final loss is

$$L = CE(M; \hat{M}) \ + \ \alpha R_L - \beta C_S(R_K), \tag{7}$$

where $\alpha, \beta \in [0, 1]$ are weight factors to balance different terms. The above loss is used to optimize network parameters θ_A and θ_B.

For the eavesdropping channel, Eve tries to decode the received Z by channel decoding and semantic decoding to obtain $M^{'}$. The reconstruction loss is

$$L_{Eve} = CE(M; M'), \tag{8}$$

similar to the legitimate channel. We let the error propagate through θ_E and θ_A such that the decoding loss of Eve could affect how Alice encodes and "encrypts" the input.

Both channels are trained end-to-end and alternately. Since we study semantic communication for natural language in this work, it is typical for all three parties to hold common prior knowledge, i.e., the vocabulary dictionary, indicated by input/output embedding blocks in Fig. 1.

III. Evaluation

A. Setup

Dataset: The English text of the European Parliament Proceedings serves as our dataset [17]. Sentences of length ranging from 4 to 30 words are selected, totaling 74,000 sentences containing 1.5 million words.

Hyper-parameters: The network structures of our system are provided in Table. I. Our Transformer-based network is trained with the Adam optimizer with a learning rate 1×10^{-4}, The variance of Gaussian noise is 0.1. The mutual information estimation network T is implemented by MLP, and is trained with Adam optimizer with a learning rate 5×10^{-5}. The variance of Gaussian noise is 0.5. By default, the row shuffling position is at 1 in Fig. 1. The weight factors α and β are set to 0.01.

Baselines: (1) *Traditional scheme:* The source coding uses Huffman coding [18], and the channel coding employs Polar coding [19] with code rate $\frac{1}{2}$. The digital modulation scheme is quadrature amplitude modulation (QAM), and one-time pad serves as the shared key. (2) *Adversarial learning (adv)*: Luo et al. [20] adopt an adversarial encryption training scheme to guarantee the accuracy of semantic communication. Since their code is not open-sourced, we re-implement the algorithm in [20].

TABLE I: The setting of the networks.

	Layer Name	Units	Activation
Sender	Embedding	128	None
	4×Transformer Encoder	128 (8 heads)	Linear
	2×Dense	128-16	Relu
Channel	Rayleigh/AWGN	None	None
Receiver and Eavesdropper	2×Dense	16-128	Relu
	4×Transformer Decoder	128 (8 heads)	Linear
	Prediction Layer	Dictionary Size	Softmax
T	3×Dense	16-100-1	Relu

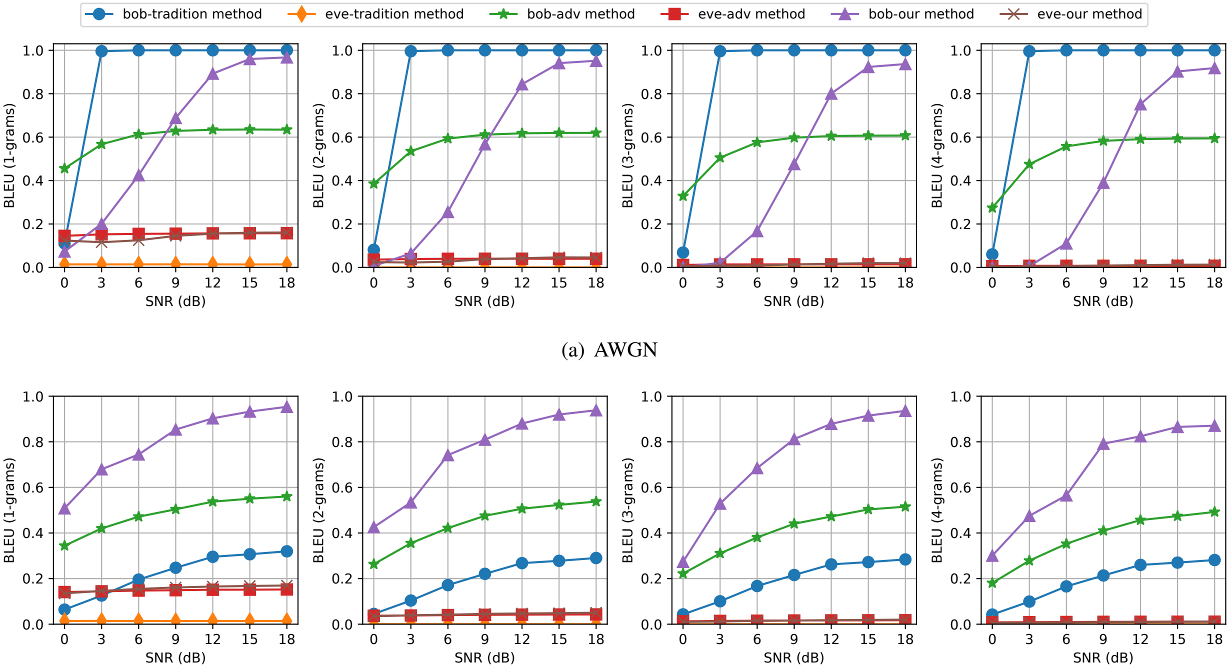
Metric: We measure the semantic error by the Bilingual Evaluation Understudy (BLEU) [21] score of the original sentence and the recovered sentence, an analogy to the bit error in physical layer communication. The higher BLEU of Bob's indicates a better transmission performance while a lower BLEU of Eve's suggests less information leakage. Since BLEU compares at n-grams, i.e., a word group, to evaluate semantic error at a larger scale, we also adopt sentence similarity [22] in NLP as the metric.

B. Results

Fig. 2(a) and 2(b) illustrate the BLEU between Alice and Bob, and the BLEU between Alice and Eve under different signal-to-noise ratios (SNRs) in Additive White Gaussian Noise (AWGN) and Rayleigh channels. The traditional method performs best in AWGN but is poor in Rayleigh channels. This is because the method aims to precisely recover every bit, suitable for AWGN channels. In contrast, our method achieves relatively high Bob's BLEU in both channels, particularly surpassing baselines by a large margin in the Rayleigh channel, which demonstrates robustness against the fading channel. Adv method has poor performance in both channels, mostly due to the difficulty in learning encryption with neural networks.

Fig. 2 also presents Eve's BLEU score under different SNRs and channels. As the traditional method employs the one-time pad which affords perfect secrecy, the score of Eve is the lowest among all. Our method and adv method share similar, close-to-zero results except for the 1-grams, demonstrating a strong capability of hiding semantic information from Eve. As we analyze, Eve's BLEU 1-grams is higher mostly because there are many single-character words in the dataset, contributing to high 1-grams scores.

Fig. 3 displays the sentence similarity scores of three methods. The results mostly agree with that under BLEU. In particular, our method achieves a sentence similarity score of over 0.8 at SNR greater than 6 dB in Rayleigh channel, indicating a successful recovery of semantic information at the sentence level. For a more intuitive understanding of our results, we sample some recovered sentences in Table. II.



(a) AWGN

(b) Rayleigh

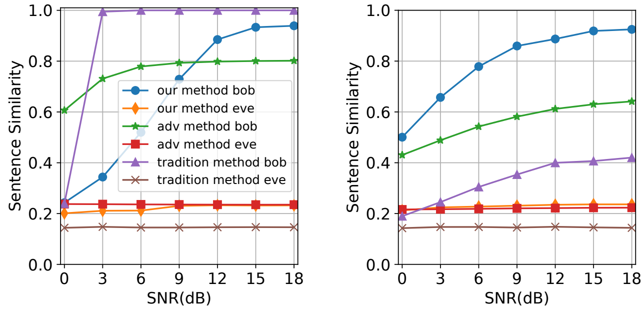
Fig. 2: Bob and Eve's BLEU scores for the three methods in (a) AWGN channel, and (b) Rayleigh fading channel.

For deeper analysis, we depict the magnitude of information leakage rate by mutual information R_L (Eq. (1)) in Fig. 4. It should be noted that the maximum mutual information here is 1 bit/channel use, and thus the results verify that our method is able to achieve sufficiently low information leakage.

C. Ablation Study

This section studies the impact of different row shuffling positions on the system. Fig. 5 shows Bob's and Eve's BLEU (1-grams) scores for the three positions. Bob achieves over 0.9 BLEU score in all three at the high SNR regime, verifying the three shuffling positions are almost equivalent to the legitimate receiver. At the low SNR regime, the performance of position 3 is comparatively poorer than position 2 which is worse than position 1. For Eve, all three positions obtain a relatively low BLEU score across different SNRs, indicating a negligible leakage to the eavesdropper. In particular, position 3 leaks the least amount of information.

The above phenomena show that placing row shuffling in position 3 reduces the reconstruction accuracy of both Bob and Eve, i.e., a degraded transmission performance at less information leakage. This may be because, at position 3, row shuffling is performed to the output of the sender's joint source-channel coding, so that only the lower part of the network (from position 3 to final loss) tries to adapt to the shuffling, with the upper part even not aware of shuffling. In contrast, if the shuffling takes place at position 1, the entire network would involve in the training to adapt to shuffling, and with more parameters, the decoding accuracy is higher. Particularly Eve, without the shared key, needs to learn decryption to random shuffling along with decoding, which is a much more difficult



(a) AWGN

(b) Rayleigh Fading

Fig. 3: Bob and Eves sentence similarity for the three methods in (a) AWGN channel, and (b) Rayleigh fading channel.

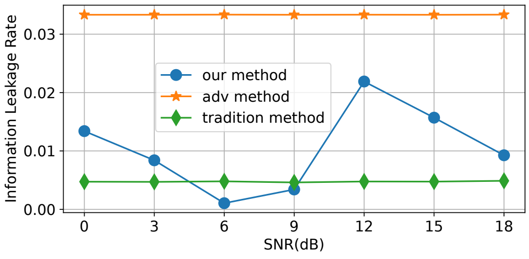


Fig. 4: Information Leakage Rates of three methods (AWGN).

TABLE II: The sentences decoded by Bob and Eve in AWGN channel, SNR= 15dB.

Alice's message	i hope that the report will have a good result and will help everyone without exceptions
Bob's decoding	i hope that the report will have a good result and will help everyone without exceptions
Eve's decoding	the next item is the commission statement on the situation in the middle east

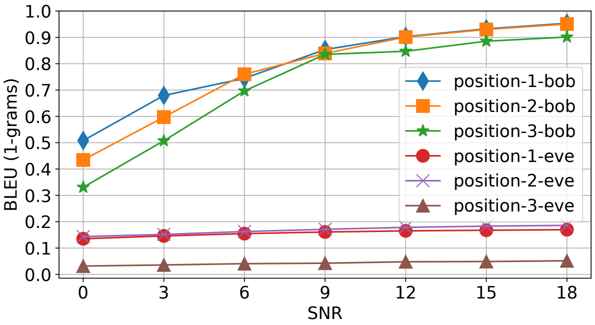


Fig. 5: The row shuffling positions have mild impacts on Bob's and Eve's BLEU scores.

task. Hence the gap between positions 1/2 and 3 is larger.

Upon a closer inspection to Bob, at SNR smaller than 6dB, position 1 has better transmission performance than 2. It may be attributed to that the semantic coding network automatically adjusts to fit shuffling at position 1. In better channel conditions, it may only take the channel encoder-decoder to fit the noisy channel, rendering almost the same performance for shuffling at positions 1 and 2. Nevertheless, all three shuffling positions are favorable depending on the specific accuracy-leakage tradeoff required.

D. Discussion

Through the above analysis, we can conclude that encryption with shuffling is a kind of destruction of the original data distribution. If the shuffling is sufficiently random, there would be no correlation between the data before and after shuffling, which almost achieves perfect secrecy. However, we also want the destruction to be totally reversible, otherwise, it would cause information loss. Taking advantage of the network structure, we are able to achieve theoretically reversible destruction, which is more explicit than the adversarial encryption training method. Thus our method ends up with a better semantic transmission performance with minimal leakage.

Even though we have taken a move, the tradeoff between communication reliability and security is yet to explore. The encoding network of Alice is required to achieve two seemingly contradictory goals: a higher transmission rate and a lower information leakage rate. How to manipulate the encoding scheme (or the destruction of data distribution) to balance the two is an interesting direction to investigate in the future.

IV. Conclusion

We propose a secure semantic communication framework that novelly utilizes random shuffling to the intermediate representation of DNN as the shared secret key for encryption. The system aims at maximizing the transmission capacity between legitimate communication parties while suppressing the leakage to the eavesdropper. Experimental results indicate that our proposed method outperforms traditional methods in terms of robustness, transmission accuracy, and security.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J*, vol. 27, no. 3, pp. 379–423, 1948.

[2] ——, "Communication theory of secrecy systems," *Bell Syst. Tech. J*, vol. 28, no. 4, pp. 656–715, 1949.

[3] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J*, vol. 54, no. 8, pp. 1355–1387, 1975.

[4] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[5] E. Ardestanizadeh, M. Franceschetti, T. Javidi, and Y.-H. Kim, "Wiretap channel with secure rate-limited feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5353–5361, 2009.

[6] H. Wang, X. Tao, N. Li, and Z. Han, "Polar coding for the wiretap channel with shared key," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 6, pp. 1351–1360, 2018.

[7] D. Chen, N. Zhang, R. Lu, X. Fang, K. Zhang, Z. Qin, and X. Shen, "An ldpc code based physical layer message authentication scheme with prefect security," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 748–761, 2018.

[8] T.-Y. Tung and D. Gunduz, "Deep joint source-channel and encryption coding: Secure semantic communications," *arXiv preprint arXiv:2208.09245*, 2022.

[9] 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," *arXiv preprint arXiv:2304.10147*, 2023.

[10] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2018, pp. 2326–2330.

[11] 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, 2021.

[12] S. Wang, J. Dai, Z. Liang, K. Niu, Z. Si, C. Dong, X. Qin, and P. Zhang, "Wireless deep video semantic transmission," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, pp. 214–229, 2023.

[13] M. Abadi and D. G. Andersen, "Learning to protect communications with adversarial neural cryptography," *arXiv preprint arXiv:1610.06918*, 2016.

[14] H. Xu, L. Xiang, H. Ye, D. Yao, P. Chu, and B. Li, "Shuffled transformer for privacy-preserving split learning," *arXiv preprint arXiv:2304.07735*, 2023.

[15] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.

[16] M. I. Belghazi, A. Baratin, S. Rajeshwar, S. Ozair, Y. Bengio, A. Courville, and D. Hjelm, "Mutual information neural estimation," in *Proc. Int. Conf. Mach. Learn.* PMLR, 2018, pp. 531–540.

[17] P. Koehn, "Europarl: A parallel corpus for statistical machine translation," in *Proceedings of machine translation summit x: papers*, 2005, pp. 79–86.

[18] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, no. 9, pp. 1098–1101, 1952.

[19] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2008, pp. 1173–1177.

[20] X. Luo, Z. Chen, M. Tao, and F. Yang, "Encrypted semantic communication using adversarial training for privacy preserving," *IEEE Commun. Lett.*, pp. 1–1, 2023.

[21] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proc. Annu. Meeting Assoc.Comput. Linguistics, Philadelphia*, 2002, pp. 311–318.

[22] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv preprint arXiv:1908.10084*, 2019.