Does Semantic Similarity Provide Enhanced Physical Layer Security?

Nihan Arı, Leila Musavian and Nikolaos Thomos

***Abstract*—This paper investigates the security of semantic communication systems over Rayleigh fading wiretap channels. Unlike traditional systems, semantic communication focuses on recovering the intended meaning rather than exact bits. To cope with this inconsistency, we propose, for the first time, using the semantic similarity metric to enhance physical layer security within a realistic framework in which noise affects the transmitted messages at the bit level similarly to traditional systems. Semantic sets are defined via semantic (cosine) similarity thresholds and decoding performance is evaluated accordingly. To address complexity challenges, we segment the transmitted messages into chunks of decreasing semantic importance. We compare two strategies for the considered semantic similarity thresholds: (a) keeping them constant across all chunks and (b) linearly scaling them to adjust their value according to the importance. Results reveal a trade-off among semantic reliability, security, and complexity. Further, our results demonstrate that by adjusting the semantic similarity threshold, the size of the semantic set is controlled, which in turn affects both decoding success and vulnerability to leakage. Finally, we note that higher thresholds improve security, but reduce reliability, while lower thresholds enhance robustness at the cost of increased risk of semantic leakage.**

***Index Terms*—Semantic communication, physical layer security**

I. Introduction

THE development of 6G networks brings a fundamental shift in communication systems by prioritizing the delivery of meaningful and purposeful content over the transmission of information. This new paradigm, known as semantic communication (SemCom), emphasizes transmitting information that is semantically important and aligns with its intended use. However, this evolution introduces new security challenges. Traditional communication models, which focus solely on delivering bit sequences, often overlook the context, content, and intent of messages. As a result, the systems become vulnerable to attackers who may exploit the conveyed semantics rather than the raw data itself. Thus, it is important to empower semantic communication systems with robust security frameworks to ensure the success of smarter and safer 6G and beyond networks [1].

N. Arı, L. Musavian, and N. Thomos are with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK (emails: {nkaraca, leila.musavian, nthomos}@essex.ac.uk).

This work was funded by the Engineering and Physical Science Research Council, UK, [Grant Numbers: EP/X04047X/1 and EP/Y037243/1].

To address these challenges, physical layer security (PLS) is a promising approach that leverages the intrinsic properties of wireless communication channels. Integrating semantic communication with PLS introduces both opportunities and challenges. While SemCom enhances communication efficiency, ensuring the security of semantic content, it demands novel strategies. This is because conventional techniques primarily focus on protecting bit sequences and ensuring their accurate reconstruction. As highlighted in [2], new PLS metrics with semantic goals are crucial for improving resilience against eavesdropping and other threats. Similarly, the work in [3] underscores the need for tailored frameworks to address the unique security risks posed by semantic systems. Traditional bit error-based security methods are no longer sufficient to capture the risks introduced by semantic communications systems. Adversaries may extract meaningful information even without perfectly decoding the transmitted bits. This limitation highlights the need for semantic-aware security frameworks that go beyond classical error metrics to quantify the risk of semantic leakage. Such frameworks should be designed to ensure reliable communication for legitimate users while effectively limiting the information gain of eavesdroppers.

Unlike most existing studies, our work focuses on a physical layer semantic communication system, where noise effect is investigated at the wireless channel itself rather than considering semantic noise. This enables the study of semantic physical layer security in a more realistic setting. Our approach offers a simple way to analyze semantic secrecy over physical channels.

A. Related Works and Motivation

Early research in [4] laid the groundwork for semantic communication by developing a theoretical framework that links semantic models to Shannon's classical information theory (CIT). This work demonstrates that Shannon's source and channel coding theorems have semantic counterparts. However, while these theorems confirm the existence of semantic coding algorithms, they offer little guidance for designing optimal ones. Building on these foundations, the work in [5] conceptualizes semantic communication as an evolution beyond Shannon's theory, particularly within the context

of Artificial Intelligence and Internet of Things (IoT) systems. The study introduces frameworks for human-to-human, human-to-machine, and machine-to-machine communication by using neural networks and distributed learning to improve 6G networks. The work in [6] explores language exploitation through joint source-channel coding and an end-to-end distortion metric and defines the semantic distortion cost region. The work in [7] establishes a systematic framework for Semantic Information Theory (SIT) by introducing synonymous mapping and metrics such as semantic entropy, mutual information, and semantic capacity. Although these efforts highlight SIT as a natural extension of CIT, a universal consensus on its formal definition has not yet emerged.

Research on semantic physical layer security is still in its early stages, but there are notable findings. In [8], semantic security is examined in type II wiretap channels (WTC II). This work finds that the secrecy capacity in this setting matches the semantic security capacity by providing robustness against strong eavesdroppers. The paper in [9] explores traditional security approaches in the Semantic IoT (SIoT), by proposing novel performance indicators, such as the probability of semantic secrecy outage and the probability of detection failure, to better assess the security of SIoT. The study also identifies semantic-level threats and the corresponding defense mechanisms. In [10], the authors propose DeepSSC, a secure semantic communication system implemented using a framework based on a deep neural network (DNN). DeepSSC is trained to balance accurate semantic recovery for legitimate users and minimize semantic leakage to eavesdroppers. BLEU score, commonly used in natural language processing (NLP), is used as a metric and the results show significant gains in terms of semantic security at high SNRs. The work in [11] presents a framework in which the transmitter sends a bitstream comprised of both traditional and semantic bit information. It uses the semantic signal as artificial noise to disrupt eavesdroppers and achieves significant gains in ergodic secrecy rate over fading wiretap channels, by optimizing transmit power, power split between semantic and bit streams, and the decoding order at the transmitter. Lastly, a recent study [12] proposes a PLS framework for integrated sensing and semantic communication systems, where a semantic base station simultaneously serves multiple users and senses a malicious target in the presence of an eavesdropper. To enhance semantic security, it optimizes beamforming for both communication and sensing, and show the trade-off between semantic secrecy rate (SSR) and sensing accuracy. While the above systems represent important advances toward achieving semantic security, they are primarily inspired by NLP and often rely on task-specific metrics such as BLEU, which limits their generality.

B. Contribution

This work investigates the security of semantic communication systems over a Rayleigh fading wiretap channel by analyzing the semantic success probabilities of legitimate users and eavesdroppers under different semantic similarity thresholds, codeword lengths and SNR values. The main contributions are as follows:

- we develop a framework to quantify semantic success and leakage probabilities using cosine similarity, where the legitimate receiver relies on threshold-based semantic sets and the eavesdropper, unaware of the threshold, performs best-match decoding;
- we offer a new perspective on evaluating physical-layer security in semantic communication systems by analyzing how semantic similarity thresholds, codeword length, and channel noise affect the trade-off between reliability and leakage;
- we introduce a chunk-wise decoding strategy that enables scalable semantic decoding for longer messages. Thresholds per chunk are adjusted by keeping them constant or linearly scaled them.

II. System Model and Problem Definition

Semantic communication redefines the objective of transmission from exact bit recovery to the preservation of meaning. While PLS traditionally aims to minimize bit error rates for legitimate users, semantic communication systems necessitate new evaluation metrics that capture how effectively meaning is conveyed and reconstructed. In this work, we adopt cosine similarity [13] as a performance metric to quantify the degree of semantic alignment between transmitted and decoded messages.

A. System Overview

Consider a Rayleigh fading wiretap channel where a transmitter (Alice) sends a binary message $M \in \{0,1\}^n$ to a legitimate receiver (Bob), while an eavesdropper (Eve) attempts to intercept as illustrated in Fig. 1. Each bit in the message is modulated by Binary Phase Shift Keying (BPSK), where the modulated signal becomes $\mathbf{x} \in \{-1,+1\}^n$, and then is transmitted through a Rayleigh fading channel with additive Gaussian noise:

$$\mathbf{y} = \mathbf{h}\mathbf{x} + \mathbf{n} \tag{1}$$

where, $\mathbf{h} = |\mathbf{g}|$, with $\mathbf{g} \sim \mathcal{CN}(0,1)$, and $\mathbf{n} \sim \mathcal{N}(0,\sigma^2)$.

B. Semantic Similarity and Semantic Sets

We determine whether a received signal conveys the same meaning by comparing the direction of vectors in the modulated multidimensional signal space. Semantic

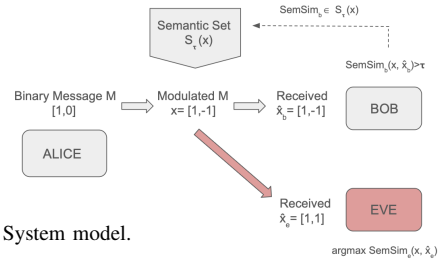

Fig. 1. System model.

accuracy is evaluated using cosine similarity between the transmitted signal $\mathbf{x}$ and a candidate $\hat{\mathbf{x}} \in \{-1,+1\}^n$:

$$\text{SemSim}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{\mathbf{x} \cdot \hat{\mathbf{x}}}{\|\mathbf{x}\| \cdot \|\hat{\mathbf{x}}\|} \tag{2}$$

Here, $(\cdot)$ denotes the dot product of two vectors, and $\|\cdot\|$ represents the Euclidean norm of a vector. A semantic set is defined for a similarity threshold $\tau \in [-1, 1]$ as:

$$\mathcal{S}_\tau(\mathbf{x}) = \{\hat{\mathbf{x}} \in \{-1,+1\}^n \,:\, \text{SemSim}(\mathbf{x}, \hat{\mathbf{x}}) \geq \tau\} \tag{3}$$

Higher τ values correspond to stricter semantic equivalence. This definition of semantic sets offers a geometric realization of the synonym mapping concept [7], where codewords achieving a similarity threshold have the same meaning. Note that, higher values of τ imply stricter semantic equivalence and smaller synonym sets.

C. Semantic Decoding and Leakage Detection

In our setting, Bob is a semantic receiver aware of the semantic similarity threshold τ, and based on this constructs the semantic set. Semantic decoding at Bob is successful if the received message belongs to this set. Additionally, Bob assigns the intended meaning (seed) to a single codeword in each semantic set. Conversely, the eavesdropper Eve, who is also a semantic receiver, performs semantic decoding and selects the codeword with the highest cosine similarity to her received signal. However, unlike Bob, she is not aware of the threshold τ and cannot construct or interpret the semantic set. Semantic leakage occurs if Eve's decoded codeword coincides with the true meaning assigned within Bob's semantic set. The full decoding and leakage detection process is summarized in Algorithm 1.

D. Chunk-Wise Similarity Threshold Scaling

The semantic set grows exponentially with the number of codeword bits. To cope with this problem, we segment the original message x into chunks of l bits. We found 8-bit chunks offer a balance between decoding complexity and allow exact computation of local semantic sets. Further, this segmentation of messages supports flexible threshold scaling across the message. This is needed as the first chunks carry more critical semantic content

Algorithm 1 Semantic Decoding and Leakage Detection

Input: Blocklength n, packet size $p = \min\{8, n\}$, packets $K = \lceil n/p \rceil$, threshold τ_{base}, decrement α, trials T
Partition indices $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$ from right to left
Set thresholds: $\tau_k = \tau_{\text{base}} - (k-1) \cdot \alpha$ for $k = 1, \ldots, K$
Generate codebook $\mathcal{C} = \{0,1\}^n$ and modulate to $\mathcal{X} = \{-1,+1\}^n$
for $i = 1$ to T **do**
  Pick random $\mathbf{x} \in \mathcal{X}$
  Semantic Seeds:
  for $k = 1$ to K **do**
    $\mathbf{x}_k = \mathbf{x}_{\mathcal{I}_k}$, define
    $$\mathcal{S}_k = \{\mathbf{z} \in \{\pm 1\}^p : \text{SemSim}(\mathbf{x}_k, \mathbf{z}) \geq \tau_k\}$$
    if $\mathcal{S}_k = \varnothing$ **then continue**
    end if
    Pick random $\mathbf{s}_k \in \mathcal{S}_k$
  end for
  Full seed: $\mathbf{s} = [\mathbf{s}_1, \ldots, \mathbf{s}_K]$
  Bob: Observe $\mathbf{y}_b = \mathbf{h}_b \mathbf{x} + \mathbf{n}_b$
  for $k = 1$ to K **do**
    $\hat{\mathbf{x}}_{b,k} = \arg\max_{\mathbf{z} \in \{\pm 1\}^p} \text{SemSim}(\text{Re}(\mathbf{y}_{b,\mathcal{I}_k}), \mathbf{z})$
  end for
  $\hat{\mathbf{x}}_b = [\hat{\mathbf{x}}_{b,1}, \ldots, \hat{\mathbf{x}}_{b,K}]$
  if $\hat{\mathbf{x}}_{b,k} \in \mathcal{S}_k \ \forall k$ **then** $T^{\text{sem}}_{\text{Bob}} \leftarrow T^{\text{sem}}_{\text{Bob}} + 1$
  end if
  Eve: Observe $\mathbf{y}_e = \mathbf{h}_e \mathbf{x} + \mathbf{n}_e$
  for $k = 1$ to K **do**
    $\hat{\mathbf{x}}_{e,k} = \arg\max_{\mathbf{z} \in \{\pm 1\}^p} \text{SemSim}(\text{Re}(\mathbf{y}_{e,\mathcal{I}_k}), \mathbf{z})$
  end for
  $\hat{\mathbf{x}}_e = [\hat{\mathbf{x}}_{e,1}, \ldots, \hat{\mathbf{x}}_{e,K}]$
  if $\hat{\mathbf{x}}_e = \mathbf{s}$ **then** $T^{\text{leak}}_{\text{Eve}} \leftarrow T^{\text{leak}}_{\text{Eve}} + 1$
  end if
end for
Compute: $\text{Pr}_{\text{B}} = T^{\text{sem}}_{\text{Bob}}/T, \quad \text{Pr}_{\text{leak}} = T^{\text{leak}}_{\text{Eve}}/T,$

compared to the next ones. Hence, a higher similarity threshold τ_{base} should be used for the first chunks to ensure accurate decoding, which can then decrease to account for the lower semantic importance of later chunks. Here, we smoothly vary the cosine similarity threshold across chunks by applying linear scaling, but other approaches may result in further improvements.

$$\tau_k = \tau_{\text{base}} - (k-1) \cdot \alpha \tag{4}$$

We start from a base threshold $\tau_{\text{base}} \in [-1, 1]$, and then decrease it by a fixed amount $\alpha > 0$ for each chunk $k = 1, 2, \ldots, K$. This generates a sequence of semantic thresholds where earlier chunks use stricter values and later chunks allow looser semantic matches. For convenience of notation, we use τ and τ_{base} interchangeably to refer to the initial base threshold.

E. Other security metrics

Apart from semantic similarity, in this section we define the following metrics for performance evaluation:

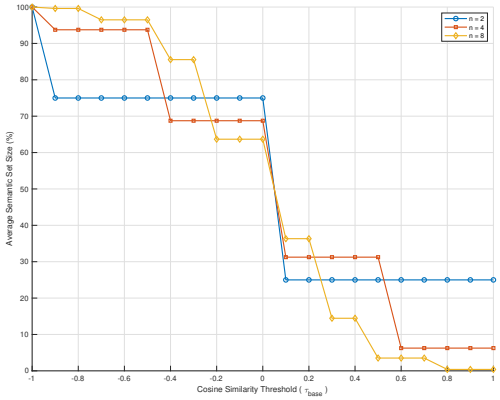

Fig. 2. Average semantic set size as a percentage of the total codebook (2^n) for different codeword lengths n, plotted against the cosine similarity threshold τ_{base}.

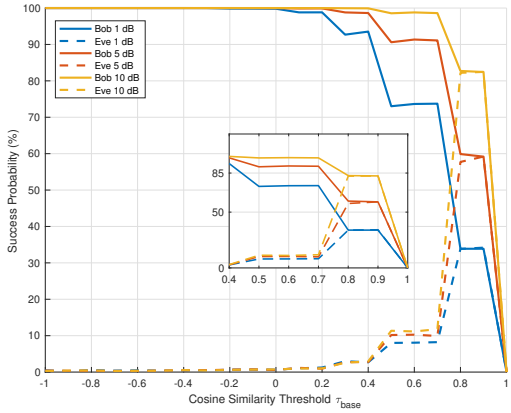

Fig. 3. Success and leakage probability as a function of the cosine similarity threshold τ_{base} for Bob and Eve under varying SNR levels for codeword length $n = 8$.

- **Semantic success** - $\text{Pr}_{\text{B}}(\tau_{\text{base}}, \text{SNR}_{\text{B}})$: It is defined as the probability that each decoded chunk lies within its corresponding semantic set. When the chunk approach is followed, the semantic success is calculated as the product of per chunk success, as to recover the actual meaning requires all the chunks to be correctly and semantically decoded.
- **Semantic leakage** - $\text{Pr}_{\text{leak}}(\text{SNR}_{\text{E}})$: It is defined as the probability that Eve's decoded chunks all match the designated seeds selected from their respective semantic sets. It is again calculated as the product each chunk matches the corresponding seed.

III. Numerical Results

Fig. 2 shows how the average semantic set size, expressed as a percentage of the total number of codewords, varies with the cosine similarity threshold τ_{base}. At lower thresholds (e.g., $\tau_{\text{base}} < 0$), the condition is loose, so almost all modulated vectors fall within each semantic set. At $\tau_{\text{base}} = -1$, every codeword is included, yielding 100% semantic set size. As τ_{base} increases, the requirement for directional similarity becomes stricter, and results in smaller semantic sets. For $\tau_{\text{base}} = 1$, the semantic set contains a single codeword, i.e., only the transmitted codeword. Semantic success then requires the decoder to recover the exact codeword. The effect of dimensionality is also evident.

When $n = 2$, the sizes of the semantic sets remain small and relatively stable across a wide range of τ_{base}, as the codewords point in distinctly different directions. In contrast, for $n = 8$, the sets are significantly larger at small τ_{base}, as codewords are more concentrated and even small angles include many nearby vectors. Moreover, some codewords have fewer semantically similar neighbors and provide greater inherent security. Assigning critical messages to these codewords may enhance secrecy but can reduce robustness, which illustrates a trade-off between reliability and leakage.

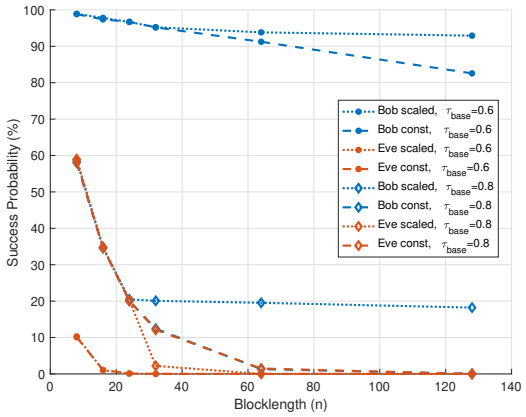

Fig. 4. Semantic decoding success of Bob and Eve vs. blocklength n, using 8-bit chunks following constant thresholds and scaled thresholds.

Fig. 3 shows how the semantic threshold τ_{base} affects the semantic success of Bob and Eve for $n = 8$. In this setting, we assume a single chunk. As τ_{base} increases, the semantic set becomes smaller and only codewords with higher cosine similarity to the transmitted message are accepted. We note that Bob performs well for low τ_{base}, but his success drops sharply beyond $\tau_{\text{base}} \approx 0.7$, especially under stronger noise. On the other hand, Eve does not know τ_{base} value, and hence, always picks the most similar codeword. From this figure, we can see Eve's success is low when semantic sets are big (low τ_{base}), because it is less likely that the correct seed will be recovered. However, when semantic set size becomes smaller (high τ_{base}), the probability of recovering the correct seed by random guessing increases. Moderate values, i.e. $\tau_{\text{base}} \in (0.5 - 0.7)$, show a trade-off on reliability and secrecy.

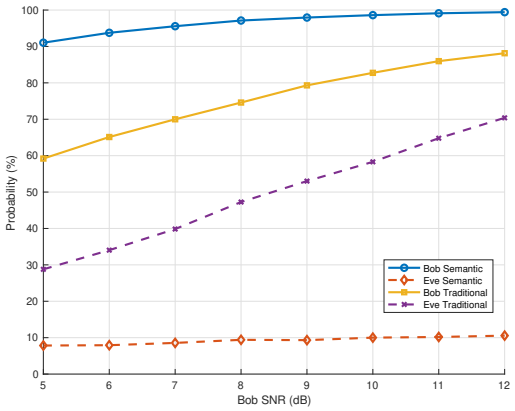

Fig. 5. Success probability of Bob and Eve under semantic and traditional decoding frameworks as a function of Bob's SNR. The blocklength is fixed at $n = 8$, and Eve's SNR is 5 dB lower than Bob's. A cosine similarity threshold of $\tau_{\text{base}} = 0.6$ is assumed.

In Fig. 4, we examine semantic success and leakage using the proposed chunk-based approach for various codeword blocklengths when 8-bit chunks are assumed. We compare τ adaptation strategies: constant thresholds, where each chunk uses the same τ_{base}, and scaled thresholds, where τ_{base} decreases linearly across chunks ($\alpha = 0.02$). Recall, a higher τ_{base} yields smaller semantic sets, while lower τ_{base} values produce larger ones. We note that Eve has a higher chance of finding the seed when the set is small, i.e., for short codewords, with $\tau_{\text{base}} = 0.8$, but her success drops to zero for larger codewords. Bob benefits from the scaled strategy, which gradually decreases the threshold across the codeword. This compensates for the impact of noise in later chunks and allows Bob to maintain higher semantic success than with a constant threshold. Overall, scaling provides a better balance between robustness and secrecy.

In Fig. 5, we compare the semantic and traditional decoding success rates of Bob and Eve over a range of SNR values to understand the potential gains coming from the use of semantic similarity. We consider that Eve faces 5 dB worse channel than Bob. The results show that semantic decoding yields consistently higher success rates for Bob compared to the traditional bit-level communication. This performance gap is more pronounced at moderate SNR values. For example, at 8 dB, Bob achieves over 95% semantic success probability, while his traditional bit-level decoding variant remains below 80%. In contrast, Eve's semantic success probability is very low and becomes close to zero even for a small blocklength of 32 bits. This is because without access to the similarity threshold τ_{base}, she cannot recover the correct seed with high probability. Differently, when Eve employs a traditional bit-level decoder, her success improves gradually with SNR as expected.

IV. Conclusion

This paper examined semantic communication over a Rayleigh fading channel, where decoding is based on cosine similarity between modulated vectors. We introduced the concept of a semantic set and proposed a framework to evaluate semantic decoding success and leakage probabilities. Simulation results reveal several important trends. First, larger codewords enhance secrecy by reducing Eve's chance of correctly guessing the transmitted signal. Second, the similarity threshold τ plays a critical role. While higher thresholds reduce semantic leakage, they may impair Bob's success under noise. Lastly, we adopt a chunk-based semantic decoding strategy to manage complexity and avoid full codebook search. This work marks an initial step toward secure semantic communication systems by highlighting the need to balance security, decoding complexity, and performance through the geometric properties of modulated signals.

References

[1] D. Gündüz, 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 JSAC*, vol. 41, no. 1, pp. 5–41, 2023.

[2] M. Mitev, A. Chorti, H. V. Poor, and G. P. Fettweis, "What Physical Layer Security Can Do for 6G Security," *IEEE OJVT*, vol. 4, pp. 375–388, 2023.

[3] Z. Yang, M. Chen, G. Li, Y. Yang, and Z. Zhang, "Secure Semantic Communications: Fundamentals and Challenges," *IEEE Netw.*, vol. 38, no. 6, pp. 513–520, 2024.

[4] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards A Theory of Semantic Communication," in *IEEE TNSE*, 2011, pp. 110–117.

[5] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, and K. Huang, "What is Semantic Communication? A View on Conveying Meaning in the Era of Machine Intelligence," *IEEE JCN*, vol. 6, no. 4, pp. 336–371, 2021.

[6] Y. Shao, Q. Cao, and D. Gündüz, "A Theory of Semantic Communication," *IEEE Trans. Mob. Comput.*, vol. 23, no. 12, pp. 12 211–12 228, 2024.

[7] K. Niu and P. Zhang, "A Mathematical Theory of Semantic Communication," *Journal on Communications*, vol. 45, no. 6, pp. 7–59, Jun. 2024.

[8] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Semantic-Security Capacity for the Physical Layer via Information Theory," in *IEEE SWSTE*, 2016, pp. 17–27.

[9] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, M. Guizani, and D. I. Kim, "Rethinking Wireless Communication Security in Semantic Internet of Things," *IEEE Wirel. Commun.*, vol. 30, no. 3, pp. 36–43, 2023.

[10] Y. Li, Z. Shi, H. Hu, Y. Fu, H. Wang, and H. Lei, "Secure Semantic Communications: From Perspective of Physical Layer Security," *IEEE Commun. Lett.*, vol. 28, no. 10, pp. 2243–2247, 2024.

[11] X. Mu and Y. Liu, "Semantic Communication-Assisted Physical Layer Security Over Fading Wiretap Channels," in *IEEE ICC*, 2024, pp. 2101–2106.

[12] H. Amiriara, M. Mirmohseni, A. Elzanaty, Y. Ma, and R. Tafazolli, "A Physical Layer Security Framework for Integrated Sensing and Semantic Communication Systems," in *IEEE WCNC*, 2025, pp. 1–6.

[13] P. Sitikhu, K. Pahi, P. Thapa, and S. Shakya, "A Comparison of Semantic Similarity Methods for Maximum Human Interpretability," in *AITB*, vol. 1, 2019, pp. 1–4.