Security and Privacy Challenges in Semantic Communication Networks

Quang Tuan Do, Dongwook Won, Tung Son Do, Thanh Phung Truong, Sungrae Cho
dept. of Computer Science and Engineering
Chung-Ang University
Seoul, South Korea
{dqtuan, dwwon, tsdo, tptruong}@uclab.re.kr, srcho@cau.ac.kr

***Abstract*—Semantic communication (SemCom) has emerged as a pivotal advancement in communication systems by focusing on the transmission of task-relevant meaning rather than raw data. This paradigm shift enables efficient communication for intelligent systems but also introduces new security and privacy risks. This paper explores these risks, reviews state-of-the-art countermeasures, and identifies key challenges and future directions in ensuring secure and private semantic communication. By addressing these issues, SemCom can fulfill its potential in domains such as healthcare, IoT, and autonomous systems.**

***Index Terms*—Semantic communication, security, privacy, adversarial attacks, semantic leakage.**

I. Introduction

Semantic communication (SemCom) has transformed traditional communication paradigms by emphasizing task-oriented and meaning-based data exchange over conventional bit-level transmission. This paradigm uses artificial intelligence (AI) and deep learning (DL) to encode, transmit, and decode information tailored to a specific task, reducing redundancy and increasing efficiency [1]. For example, in a healthcare scenario, rather than sending raw imaging data, SemCom can send the semantic message "abnormal growth detected," significantly reducing bandwidth requirements while retaining actionable insights.

SemCom's deployment is widespread, spanning IoT, autonomous driving, and edge computing. In self-driving cars, semantic systems can share important information like "obstacle detected" rather than high-resolution video streams, resulting in faster and more efficient communication. Similarly, in industrial IoT, semantic communication enables smart decision-making by sending context-aware summaries of sensor data rather than raw measurements [2].

However, these advancements present unique security and privacy challenges. Unlike traditional communication systems, which focus on protecting raw data, SemCom systems face threats to their semantic representations and AI models. Adversaries can manipulate transmitted semantics to confuse decision-making processes, such as changing the classification of traffic signs in self-driving cars or introducing malicious noise into medical diagnostics. Privacy concerns are equally important, as semantic data abstractions may inadvertently reveal sensitive user information, even when encrypted or anonymized. Addressing these issues is critical to the secure and ethical deployment of SemCom systems [3].

This paper investigates these challenges in detail, analyzing the unique vulnerabilities of semantic systems and reviewing state-of-the-art solutions. It also identifies key research directions, such as the development of privacy-preserving techniques, adversarial robustness, and trust mechanisms, to foster secure and privacy-aware semantic communication.

II. Security Threats in Semantic Communication

Semantic communication introduces a shift in communication paradigms but also opens new attack surfaces that are unique to its operation. Below, we provide a detailed discussion of the major threats.

A. Semantic Leakage

One of the most dangerous risks to SemCom systems is semantic leakage. SemCom transmits higher-level abstractions or interpretations of data, which by their very nature contain contextual information, in contrast to traditional communication, where encryption of raw data is frequently adequate to protect information. Sensitive information about the task, the sender, or even the underlying data itself may inadvertently be revealed in this context [4].

For example, even when anonymization techniques are used, sending semantic messages like "Stage 3 cancer detected" in healthcare systems that use SemCom may unintentionally expose private patient information. Particularly in systems with frequent interactions, attackers could intercept these semantic representations and deduce patient conditions or patterns in historical data. In edge computing settings, where numerous devices with different degrees of trust interact, these risks are increased.

Advanced methods like differential privacy, which introduces noise into semantic outputs without compromising their usefulness for tasks downstream, are necessary to mitigate semantic leakage. Finding the ideal balance between semantic accuracy and privacy protection, however, continues to be a crucial research challenge.

B. Adversarial Attacks

Adversarial attacks pose a significant threat to AI-driven semantic systems. These attacks exploit vulnerabilities in the deep learning models used for encoding and decoding semantics [5].

- Evasion Attacks: Small, imperceptible perturbations added to inputs can lead to incorrect semantic representations. For example, an adversary could subtly manipulate a vehicle's sensor data, causing an autonomous car to misinterpret a "stop sign" as a "yield sign".
- Poisoning Attacks: By injecting malicious data during the training phase, attackers can compromise the integrity of the semantic model. Poisoned models may misclassify critical inputs, potentially leading to catastrophic failures in applications like industrial IoT or healthcare.
- Inference Attacks: Adversaries can also infer private details from the intermediate features of semantic models, particularly in shared learning environments like federated learning.

The necessity of adversarial robustness in semantic systems is highlighted by these attacks. Semantic validation mechanisms, which guarantee that outputs are consistent with known constraints, and adversarial training, which exposes models to adversarial examples during training, are two techniques that have shown promise but are still resource-intensive.

C. Trust and Model Integrity

The reliability of SemCom systems depends on the integrity of shared semantic models [6]. Semantic models and shared ontologies are critical in multi-agent environments like smart cities and industrial IoT networks for ensuring device interoperability. However, these shared resources are also attractive targets for attackers looking to manipulate or compromise system-wide communication.

For example, if an attacker tampers with a collaborative manufacturing system's shared semantic model, it may result in widespread errors, such as misinterpretation of production data or incorrect task execution. Blockchain technology has been proposed as a solution for improving trust in such environments [7]. Blockchain can prevent unauthorized changes by keeping an immutable record of model updates and verifying semantic data's provenance.

D. Contextual Inference Risks

Semantic systems rely on contextual data to improve task performance, but this reliance raises the possibility of inference attacks [8]. Adversaries can infer user behaviors, preferences, or environmental contexts by analyzing transmitted semantics patterns. For example, in smart home systems, repeated semantic messages about "low temperature detected" may reveal the household's occupancy patterns, posing physical security risks.

III. Privacy Preservation Techniques

As semantic communication becomes more prevalent in sensitive domains such as healthcare, autonomous systems, and IoT, privacy concerns grow. This section discusses various techniques for protecting privacy in SemCom, ranging from traditional encryption methods to emerging solutions such as federated learning and homomorphic encryption. Each of these techniques contributes significantly to lowering the risk of information leakage and improving SemCom system privacy guarantees.

A. Differential Privacy

Differential privacy (DP) has emerged as an effective framework for protecting privacy, particularly in the context of federated learning. In SemCom, where models are trained to derive meanings from contextual data, DP ensures that the output contains no specific information about individual data points. DP works by introducing noise into the data or model updates in such a way that the probability of any given data point being included in the dataset remains roughly constant, regardless of whether it is included or not.

For example, in a healthcare application, DP can be used to anonymize semantic representations of medical data before it is transmitted over a network. By introducing noise into the semantic message, DP ensures that an adversary who intercepts the message cannot deduce private information about the patient. However, the effectiveness of DP is dependent on the careful tuning of noise parameters in order to balance privacy with the accuracy and utility of transmitted semantics. Recent advances have attempted to increase the utility of DP in wiretap channel communication by incorporating it into more sophisticated learning models [9].

B. Homomorphic Encryption

Homomorphic encryption (HE) enables computations on encrypted data without the need to decrypt it first [3]. This property is especially useful for privacy-preserving semantic communication in cloud or edge environments, where data is processed remotely but must be kept confidential. SemCom allows encrypted semantic messages to be sent to a server, where tasks like semantic decoding or inference can be performed without exposing sensitive data.

Homomorphic encryption is particularly useful when sensitive data is being transmitted between multiple devices in a distributed system. For example, in smart cities or industrial IoT networks, nodes may send encrypted semantic data (e.g., "high temperature detected in sensor A") to a central processing unit. The central unit can then perform functions such as anomaly detection without disclosing any specific sensor data. Despite its obvious benefits, homomorphic encryption is computationally expensive, and its adoption remains limited due to the high overhead required for operations. Research is being conducted to optimize HE schemes for efficiency, particularly in real-time applications where latency and computation power are critical constraints [10].

C. Adversarial Robustness Strategies

Adversarial robustness is essential for ensuring the security and dependability of semantic communication systems against adversarial attacks. Adversarial attacks on AI-driven models for semantic inference can significantly reduce the quality of semantic communication by introducing malicious perturbations into the input data. To mitigate such threats, robust

semantic models can be trained with adversarial examples, exposing them to a wide range of possible attack scenarios.

Adversarial training involves adding small perturbations to the training data to simulate real-world attacks, thereby teaching the model to withstand such changes without significantly degrading performance. Furthermore, semantic consistency checks can be used to compare decoded semantics to expected outcomes. For example, in autonomous driving, a consistency check may ensure that an interpreted "stop sign" always corresponds to a semantic action such as "apply brakes," ensuring safety even if adversarial modifications occur.

Techniques like adversarial training have shown promise in improving model robustness, but they come at a cost in terms of computation time and model complexity. As a result, researchers are increasingly focusing on developing lightweight adversarial defense mechanisms that do not degrade the efficiency of task-oriented communication [11].

D. Federated Semantic Learning

Federated learning is a decentralized approach to training machine learning models that enables devices to collaborate on model building while keeping raw data on the local device. This approach has proven particularly useful in privacy-sensitive environments, as it eliminates the need for data to leave the local device, lowering the risk of data exposure during transmission. Federated learning in SemCom enables models to learn semantic representations without sharing sensitive raw data across devices [12].

Federated learning is especially useful in situations where large-scale collaboration is required while data privacy must be maintained. For example, in healthcare systems, hospitals can collaborate to create a semantic model for medical diagnostics without sharing patient data directly. Instead, each hospital trains a local model and only sends model updates (gradients) to a central server for aggregation. Secure aggregation improves federated learning's privacy by ensuring that the central server cannot access individual updates, preventing sensitive information from being leaked. Despite its promising benefits, there are still challenges in addressing potential vulnerabilities, such as model inversion attacks, which allow adversaries to infer private details from model updates, and the computational burden of model training on resource-constrained devices.

IV. Future Research Directions

As the use of semantic communication grows, it is critical to address emerging security and privacy concerns. The following research directions offer a road map for improving the robustness and privacy of SemCom systems in the coming years.

A. Cross-Layer Security

Semantic communication systems operate across multiple layers, including the physical, network, and application layers. A critical area of research is integrating security measures across these layers to address vulnerabilities that may arise at the intersections. For example, a physical layer security mechanism could help prevent eavesdropping or jamming, whereas network layer encryption could protect data in transit. At the semantic layer, adversarial robustness and privacy preservation can help reduce the risks associated with model manipulation and semantic leakage. A comprehensive cross-layer security framework would offer complete protection by coordinating security measures across all layers of the communication stack.=

B. Lightweight Security for IoT

The rapid growth of IoT and edge computing devices poses a significant challenge to privacy-preserving SemCom, as these devices frequently have limited computational resources and energy constraints. Developing lightweight cryptographic and privacy-preserving mechanisms that strike a balance between efficiency and privacy protection is an important area of current research. For instance, lightweight homomorphic encryption schemes and efficient privacy-preserving aggregation techniques are required to ensure privacy while not overburdening resource-constrained devices.

Another promising direction is the development of privacy-preserving protocols that require minimal computational overhead while maintaining semantic message confidentiality. Techniques such as secure multi-party computation (SMPC) and privacy-preserving federated learning can be adapted to work in resource-constrained environments, allowing for privacy guarantees in large-scale distributed networks.

C. Explainable Semantic Models

As semantic communication models become more complex, explainability (XAI) in AI-driven systems will be critical for establishing trust and ensuring that system decisions are understandable and justifiable. Stakeholders in security-critical applications such as autonomous vehicles or medical diagnostics must understand how decisions are made using transmitted semantics. Explainable semantic models can improve transparency by allowing users to validate decisions and detect potential issues or errors during the inference process.

Future research in this area could concentrate on developing explainable AI methods for semantic communication, such as providing interpretable visualizations of the reasoning underlying semantic decoding or classification. Furthermore, combining explainability with security mechanisms can help in the detection of vulnerabilities, such as adversarial manipulations, as well as improving overall system robustness.

D. Ethical and Regulatory Frameworks

As SemCom evolves, ethical considerations and regulatory frameworks will have a significant impact on its deployment, especially in sensitive fields such as healthcare, autonomous driving, and smart cities. Research in this area should concentrate on creating guidelines for data privacy, user consent, and accountability in SemCom systems. Governments and organizations must work together to establish standards that protect privacy and security while encouraging innovation.

For example, the European Union's General Data Protection Regulation (GDPR) could be used as a model for addressing

privacy concerns in semantic communication systems by ensuring that data is collected, stored, and shared in a transparent and ethical manner. There is also a need for research to establish global standards for the ethical use of SemCom, particularly in high-stakes applications that require balancing privacy and system performance.

V. Conclusion

Semantic communication has the potential to transform communication systems by moving from raw data exchange to meaning-driven communication. However, this transformation raises new security and privacy concerns, such as semantic leakage, adversarial attacks, and model trust issues. Addressing these concerns necessitates a multifaceted approach that combines traditional encryption techniques with novel solutions like federated learning, adversarial robustness, and explainable AI. By improving these privacy-preserving techniques and developing cross-layer security mechanisms, researchers can enable secure and privacy-aware deployment of SemCom in mission-critical applications, paving the way for safer, more efficient communication networks.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00453301)

VI. References Section

References

[1] W. Yang, H. Du, Z. Q. Liew, W. Y. B. 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, 2022.

[2] H. Xie and Z. Qin, "A lite distributed semantic communication system for internet of things," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 142–153, 2020.

[3] M. Shen, J. Wang, H. Du, D. Niyato, X. Tang, J. Kang, Y. Ding, and L. Zhu, "Secure semantic communications: Challenges, approaches, and opportunities," *IEEE Network*, 2023.

[4] Y. Li, Z. Shi, H. Hu, Y. Fu, H. Wang, and H. Lei, "Secure semantic communications: From perspective of physical layer security," *IEEE Communications Letters*, 2024.

[5] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, "Is semantic communication secure? a tale of multi-domain adversarial attacks," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 50–55, 2023.

[6] Y. Wang, "Semantic communication networks empowered artificial intelligence of things," in *2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT)*. IEEE, 2024, pp. 189–193.

[7] 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, 2023.

[8] Y. Wang, S. Guo, Y. Deng, H. Zhang, and Y. Fang, "Privacy-preserving task-oriented semantic communications against model inversion attacks," *IEEE Transactions on Wireless Communications*, 2024.

[9] W. Chen, S. Tang, and Q. Yang, "Enhancing image privacy in semantic communication over wiretap channels leveraging differential privacy," *arXiv preprint arXiv:2405.09234*, 2024.

[10] B. Kim, S. Heo, J. Lee, S. Jeong, Y. Lee, and H. Kim, "Compiler-assisted semantic-aware encryption for efficient and secure serverless computing," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5645–5656, 2020.

[11] G. Nan, Z. Li, J. Zhai, Q. Cui, G. Chen, X. Du, X. Zhang, X. Tao, Z. Han, and T. Q. 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, 2023.

[12] H. Wei, W. Ni, W. Xu, F. Wang, D. Niyato, and P. Zhang, "Federated semantic learning driven by information bottleneck for task-oriented communications," *IEEE Communications Letters*, 2023.