Robust and secure semantic communication methods over integrated terrestrial and non-terrestrial networks

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program
SSD: IINF-03/A–Telecommunications

Final Dissertation

Robust and Secure Semantic Communication Methods Over Integrated Terrestrial and Non-terrestrial Networks

---

by
Narimani Kenari Mahshid

Supervisors:

Prof. Nicola Cordeschi

Prof. Luigi Alfredo Grieco

Coordinator of Ph.D. Program:

Prof. Nicola Giaquinto

Politecnico di Bari

LIBERATORIA PER L'ARCHIVIAZIONE DELLA TESI DI DOTTORATO

Al Magnifico Rettore
del Politecnico di Bari

Il/la sottoscritto/a Mahshid Narimani Kenari nato/a a IRAN il 1992.08.03 residente a Bari (BA) in via Edomondo de Amicis 4 e-mail m.narimanikenari@phd.poliba.it iscritto al 3° anno di Corso di Dottorato di Ricerca in di Dottorato di ricerca in Ingegneria Elettrica e dell'Informazione ciclo 38 ed essendo stato ammesso a sostenere l'esame finale con la prevista discussione della tesi dal titolo:

Robust and Secure Semantic Communication Methods Over Integrated Terrestrial and Non-terrestrial Networks

DICHIARA

1) di essere consapevole che, ai sensi del D.P.R. n. 445 del 28.12.2000, le dichiarazioni mendaci, la falsità negli atti e l'uso di atti falsi sono puniti ai sensi del codice penale e delle Leggi speciali in materia, e che nel caso ricorressero dette ipotesi, decade fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;
2) di essere iscritto al Corso di Dottorato di ricerca in Ingegneria Elettrica e dell'Informazione ciclo 38, corso attivato ai sensi del "*Regolamento dei Corsi di Dottorato di ricerca del Politecnico di Bari"*, emanato con D.R. n.286 del 01.07.2013;
3) di essere pienamente a conoscenza delle disposizioni contenute nel predetto Regolamento in merito alla procedura di deposito, pubblicazione e autoarchiviazione della tesi di dottorato nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica;
4) di essere consapevole che attraverso l'autoarchiviazione delle tesi nell'Archivio Istituzionale ad accesso aperto alla letteratura scientifica del Politecnico di Bari (IRIS-POLIBA), l'Ateneo archivierà e renderà consultabile in rete (nel rispetto della Policy di Ateneo di cui al D.R. 642 del 13.11.2015) il testo completo della tesi di dottorato, fatta salva la possibilità di sottoscrizione di apposite licenze per le relative condizioni di utilizzo (di cui al sito http://www.creativecommons.it/Licenze), e fatte salve, altresì, le eventuali esigenze di "embargo", legate a strette considerazioni sulla tutelabilità e sfruttamento industriale/commerciale dei contenuti della tesi, da rappresentarsi mediante compilazione e sottoscrizione del modulo in calce (Richiesta di embargo);
5) che la tesi da depositare in IRIS-POLIBA, in formato digitale (PDF/A) sarà del tutto identica a quelle **consegnate**/inviate/da inviarsi ai componenti della commissione per l'esame finale e a qualsiasi altra copia depositata presso gli Uffici del Politecnico di Bari in forma cartacea o digitale, ovvero a quella da discutere in sede di esame finale, a quella da depositare, a cura dell'Ateneo, presso le Biblioteche Nazionali Centrali di Roma e Firenze e presso tutti gli Uffici competenti per legge al momento del deposito stesso, e che di conseguenza va esclusa qualsiasi responsabilità del Politecnico di Bari per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;
6) che il contenuto e l'organizzazione della tesi è opera originale realizzata dal sottoscritto e non compromette in alcun modo i diritti di terzi, ivi compresi quelli relativi alla sicurezza dei dati personali; che pertanto il Politecnico di Bari ed i suoi funzionari sono in ogni caso esenti da responsabilità di qualsivoglia natura: civile, amministrativa e penale e saranno dal sottoscritto tenuti indenni da qualsiasi richiesta o rivendicazione da parte di terzi;
7) che il contenuto della tesi non infrange in alcun modo il diritto d'Autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta.

Luogo e data Bari, 31.03.2026 Firma Mahshid Narimani Kenari

Il/La sottoscritto, con l'autoarchiviazione della propria tesi di dottorato nell'Archivio Istituzionale ad accesso aperto del Politecnico di Bari (POLIBA-IRIS), pur mantenendo su di essa tutti i diritti d'autore, morali ed economici, ai sensi della normativa vigente (Legge 633/1941 e ss.mm.ii.),

CONCEDE

- al Politecnico di Bari il permesso di trasferire l'opera su qualsiasi supporto e di convertirla in qualsiasi formato al fine di una corretta conservazione nel tempo. Il Politecnico di Bari garantisce che non verrà effettuata alcuna modifica al contenuto e alla struttura dell'opera.
- al Politecnico di Bari la possibilità di riprodurre l'opera in più di una copia per fini di sicurezza, back-up e conservazione.

Luogo e data Bari, 31.03.2026 Firma Mahshid Narimani Kenari

Department of Electrical and Information Engineering
ELECTRICAL AND INFORMATION ENGINEERING
Ph.D. Program
SSD: IINF-03/A–Telecommunications

Final Dissertation

Robust and Secure Semantic Communication Methods Over Integrated Terrestrial and Non-terrestrial Networks

by

Narimani Kenari Mahshid

Mahshid Narimani Kenari

Referees:

Prof.ssa Claudia Campolo

Prof. Simone Morosi

Supervisors:

Prof. Nicola Cordeschi

Prof. Luigi Alfredo Grieco

Coordinator of Ph.D Program:

Prof. Nicola Giaquinto

Course n°38, 01/04/2023-31/03/2026

«Power belongs to those who truly know,
Through knowledge, aged hearts in youth will glow»
*From **Ferdowsi**'s Shah-nameh*

POLITECNICO DI BARI

Abstract

Department of Electrical and Information Engineering

Doctor of Philosophy

Robust and secure semantic communication methods over integrated terrestrial and non-terrestrial networks

by Mahshid Narimani Kenari

The integration of Terrestrial Networks (TNs) and Non-Terrestrial Networks (NTNs) is expected to play a central role in future Sixth Generation (6G) systems by providing seamless connectivity across ground, air, and space segments. This evolution supports applications such as remote sensing and disaster management, which require the reliable transmission of large volumes of data under stringent bandwidth, latency, and energy constraints. However, this trend exposes communication systems to adverse channel conditions, limited resources, and security risks due to the broadcast nature of wireless transmission. To overcome these challenges, it is necessary to develop communication and security mechanisms that are not only robust and resource-efficient, but also suitable for Artificial Intelligence (AI)-native 6G architectures. This thesis addresses the challenges through three complementary contributions that span transmission efficiency, privacy, and adaptive security. First, it proposes a lightweight Semantic Communication (SemCom) framework based on deep Joint Source-Channel Coding (JSCC) for robust video transmission over noisy wireless channels. By integrating semantic attention and Signal-to-Noise Ratio (SNR)-aware modulation, the proposed method improves the transmission of task-relevant information while maintaining low computational cost. The proposed architecture makes the transmission suitable for resource-constrained scenarios relevant to TN–NTN systems. Second, the thesis investigates security vulnerabilities in SemCom, with a particular focus on eavesdropping and Model Inversion Eavesdropping Attacks (MIEAs). To mitigate these threats, it introduces a lightweight and reversible protection mechanism, Key-Assisted Protection for Deep Joint Source-Channel Coding (KAP-DJSCC), which secures the transmitted stream through key-based transformations. Third, the thesis develops a risk-based adaptive security framework for AI-native ’6G systems which provides a dynamic selection of countermeasures under resource and budget constraints. Overall, the contributions of this thesis address three important aspects of 6G systems which are applicable for TN-NTN: communication efficiency, semantic security, and adaptive cyber-risk management.

Acknowledgements

My time in the Telematics Lab was a meaningful experience, as I had the opportunity to work alongside knowledgeable and inspiring young researchers. As I write this thesis, I would first like to express my sincere gratitude to my supervisors, Prof. Nicola Cordeschi and Prof. Luigi Alfredo Grieco, for their invaluable guidance, support, and insightful advice throughout this journey. Without them, this PhD would not have been possible.

I would like to express my deepest appreciation to Giancarlo, Alessio, and Ilaria for their helpful, constructive, and kind insights, which have been of great value to me. I would also like to thank Prof. Giovanni Apruzzese, my supervisor during my research visit abroad, for his guidance and for all that he taught me. I am grateful for the friendly environment of the Telematics Lab and for all the people who were part of it. In particular, I would like to thank Federica for always being available to help me with my many questions from all directions. My thanks also go to my colleague, Gianvito for his great contributions and efforts in writing our joint paper.

Finally, I want to thank my parents, who are thousands of kilometers away from me, yet always remain close to my heart. Their love, support, and presence in my life have always been my greatest strength.

*All contributions are the result of research activities carried out by different academic and industrial partners, collaborating as a partnership in the context of the SNS JU project 6G-GOALS under the EU's Horizon program Grant Agreement No 101139232 and by European Union under the Italian National Recovery and Resilience Plan of NextGenerationEU, partnership on Telecommunications of the Future (PE00000001 - program RESTART).

Contents

List of Figures

List of Tables

List of Acronyms

5G Fifth Generation

6G Sixth Generation

AE Autoencoder

AES Advanced Encryption Standard

AI Artificial Intelligence

AIA Attribute Inference Attack

AN Artificial Noise

ANN Artificial Neural Network

APT Advanced Persistent Threat

ATN Adversarial Transformation Network

AVC Advanced Video Coding

AWGN Additive White Gaussian Noise

BER Bit Error Rate

BLEU Bilingual Evaluation Understudy

CATFL Certificateless Authentication-based Trustworthy Federated Learning

CDN Content Delivery Network

CNN Convolutional Neural Network

COR Correlation

CSI Channel State Information

DDOS Distributed Denial-of-Service

DH Diffie-Hellman

DL Deep Learning

DJSCC Deep Joint Source-Channel Coding

DNN Deep Neural Network

FaaS Firewall as a Service

FL Federated Learning

GA Genetic Algorithm

GAN Generative Adversarial Network

GDPR General Data Protection Regulation

GEO Geostation Earth Orbit

GoP Group of Pictures

HBC Honest But Curious

HAP High Altitude Platform

IB Information Bottleneck

IoT Internet of Things

IRAM2 Information Risk Assessment Methodology 2

ISAC Integrated Sensing and Communication

ISF Information Security Forum

ISSC Integrated Sensing and Semantic Communication

JSCC Joint Source-Channel Coding

KAP-DJSCC Key-Assisted Protection for Deep Joint Source-Channel Coding

KMIEA Key-assisted Model Inversion Eavesdropping Attack

KB Knowledge Base

KPI Key Performance Indicator

MEO Medium Earth Orbit

LDPC Low-Density Parity Check

LFBVS Local Feature Based Visual Security

LLR Log-Likelihood Ratio

LPIPS Learned Perceptual Image Patch Similarity

LWE Learning With Errors

LEO Low Earth Orbit

MCC Matthews Correlation Coefficient

MIEA Model Inversion Eavesdropping Attack

MILP Mixed Integer Linear Programming

MISO Multiple-Input Single-Output

ML Machine Learning

MLP Multi-Layer Perceptron

mIoU Mean Intersection over Union

mMTC massive Machine-Type Communications

MSE Mean Squared Error

MS-SSIM Multi-Scale Structural Similarity Index Measure

nnz number of non-zero elements

NT Non-Terrestrial

NTN Non-Terrestrial Network

OAR Object-Attribute-Relation

OCR Optical Character Recognition

OFDM Orthogonal Frequency Division Multiplexing

PLS Physical Layer Security

PSNR Peak Signal-to-Noise Ratio

QAM Quadrature Amplitude Modulation

RICS Reconfigurable Intelligent Computational Surface

RIS Reconfigurable Intelligent Surface

RL Reinforcement Learning

ROI Region of Interest

RSA Rivest–Shamir–Adleman

S-BLEU Semantic Bilingual Evaluation Understudy

SaaS Security-as-a-Service

SASE Secure Access Service Edge

SDWAN Software-defined Wide Area Network

SemCom Semantic Communication

SER Symbol Error Rate

SGM Segmentation Adjacency Matrix

SINR Signal-to-Interference-plus-Noise Ratio

SISO Single-Input Single-Output

SNR Signal-to-Noise Ratio

SSR Security Semantic Transmission Rate

SSIM Structural Similarity Index Measure

STAR-RIS Simultaneous Transmitting and Reflecting Reconfigurable Intelligent Surface

SWG Secure Web Gateway

TN Terrestrial Network

UAV Unmanned Aerial Vehicle

UE User Equipment

URLLC Ultra-Reliable Low Latency Communication

VPN Virtual Private Network

WAN Wide Area Network

ZTA Zero-Trust Architecture

Scientific Contributions

All the scientific contributions produced during the doctoral course are listed below.

International Journals:

- MN Kenari, G Sciddurlo, LA Grieco, and N Cordeschi, “Privacy Preserving in Wiretap Channels Beyond Shannon’s Paradigm: Current Trends and Future Challenges”, IEEE Journal of Networks and Communications, 2026.
- MN Kenari, G Sciddurlo, A Fascista, I Cianci, LA Grieco, and N Cordeschi, “Risk-based adaptive security framework for 6G Edge”, will be submitted to IEEE Transactions on Information Forensics and Security.

International Conferences:

- MN Kenari, N Cordeschi, and LA Grieco, "KAP-DJSCC: Key-Assisted Protection for Deep Joint Source-Channel Coding Under Model Inversion Eavesdropping Attacks", 2025 16th IFIP Wireless and Mobile Networking Conference (WMNC), Leuven, Belgium, 2025.
- G Coppola, MN Kenari, G Sciddurlo, N Cordeschi, LA Grieco, and G Boggia, “Semantic-Aware Attention-Driven JSCC for Efficient Video Transmission over Wireless Channels”, The Fourth DeepWireless Workshop on Deep Learning for Wireless Communications, Sensing, and Security (DeepWireless) , Tokyo, Japan, 2026.

Introduction

0.1 Background and Motivation

The expanding need for high data rates, seamless coverage, and diverse content requires a significant advancement in telecommunications [1]. Modern applications such as autonomous vehicles and smart farming need reliable, low-latency, and high-throughput communication across diverse geographical environments [2]. Despite the remarkable advancements achieved by Fifth Generation (5G) networks, for example, Ultra-Reliable Low Latency Communications (URLLCs), current terrestrial infrastructures remain insufficient to provide continuous and global coverage [3], [4]. In particular, remote, rural, maritime, and disaster-affected regions continue to experience limited or no connectivity due to the high deployment and maintenance costs associated with terrestrial communication systems, such as fiber optics and base station networks [5].

To provide a seamless and ubiquitous connectivity, 6G communication systems are envisioned to adopt a paradigm shift toward the TN and NTN. This integration enables the cooperation of ground-based infrastructures with aerial and spaceborne platforms, forming a unified space–air–ground communication architecture. NTNs include a wide range of platforms, such as Unmanned Aerial Vehicles (UAVs), High Altitude Platforms (HAPs), and satellite systems operating in Low Earth Orbit (LEO), Medium Earth Orbit (MEO), and Geostation Earth Orbit (GEO). By leveraging these complementary layers, TN–NTN integration aims to provide global coverage, enhance network resilience, and continuous service delivery in both densely populated and remote areas [1].

Beyond coverage extension, TN–NTN integration is expected to play a crucial role in emerging applications that require remote-area sensing,maritime communication, and real-time decision-making. For example, in environmental monitoring and disaster management, satellite and aerial platforms can collect large volumes of visual and sensor data from geographically dispersed regions. Similarly, in intelligent transportation systems, UAVs and satellites can assist ground networks in providing connectivity and situational awareness. These applications impose stringent requirements on communication systems, including ultra-low latency, high reliability, and efficient utilization of limited resources [1], [3].

0.2 Challenges and Problem Statement in TN–NTN Systems

Despite the significant potential, the integration of TN–NTN systems introduces a number of fundamental challenges that differentiate it from conventional terrestrial

networks:

First, visual data is expected to dominate network traffic in NTNs such as surveillance, remote sensing, and environmental monitoring. The rapid growth of multimedia data, particularly images and videos, imposes substantial demands on bandwidth and energy resources. However, TN-NTN platforms often operate under strict constraints in terms of power, latency, and bandwidth. As a result, transmitting large volumes of raw data using conventional communication techniques is inefficient in TN–NTN environments. This requires efficient mechanisms for transmitting and processing data. Extracting the most important part of the data and transmitting it is a viable solution to the limited resources in NTNs. Conventional communication paradigms, which focus primarily on accurate bit-level transmission, only consider the complete message transmission and do not optimize the resource consumption according to the importance of the information.

Second, in integrated TN–NTN systems, the wireless communication medium is inherently open and broadcast in nature, allowing transmitted signals to be received not only by intended users but also by unintended or malicious entities. This fundamental characteristic makes such systems particularly vulnerable to a range of security threats, including spoofing, jamming, and eavesdropping. Unlike active attacks, eavesdroppers can operate passively by exploiting their knowledge of communication protocols, making them difficult to detect while still posing a serious risk to data confidentiality. These vulnerabilities are further intensified in TN–NTN environments due to the large coverage areas of satellites, UAVs, and HAPs, which propagate the signal over vast geographical regions. In particular, the growing deployment of dense satellite constellations significantly expands the number of potential interception points. Satellites themselves can act as powerful eavesdroppers capable of intercepting transmissions from ground terminals without authorization. Thus, the open nature of TN–NTN systems increases the risk of eavesdropping that makes the protection of sensitive information a critical and challenging task [4], [6].

Third, the evolution toward 6G is expected to transform communication systems to AI-native architectures, where learning-based mechanisms are integrated into the physical layer, resource management, signal processing, and network orchestration [7]. While this provides intelligent and autonomous operation in highly dynamic TN–NTN environments, it also expands the attack surface beyond conventional wireless vulnerabilities. As mentioned, the open and broadcast nature of wireless channels continues to expose systems to threats such as jamming, interference, and eavesdropping, particularly in large-scale NTN deployments [4]. In parallel, the reliance on data-driven models introduces AI-specific vulnerabilities, including data poisoning, adversarial perturbations, and model inversion. More critically, the convergence of these two domains gives rise to hybrid attacks, where adversaries exploit the wireless medium and the learning process simultaneously, for example, MIEAs [8]. This interaction is particularly severe in AI-driven communication systems, where the meaning of transmitted information can be altered without necessarily affecting bit-level accuracy. Furthermore, the distributed nature of TN–NTN systems, coupled with limited onboard computation and heterogeneous data sources, makes the protection more challenging . Consequently, ensuring security in AI-native 6G networks requires a joint consideration of wireless and AI vulnerabilities, as their combined effect can amplify risks and introduce new attacks [7], [8].

These challenges highlight the need for new communication paradigms that extend beyond conventional bit-oriented designs. Future TN–NTN systems must not only efficiently handle massive volumes of data under stringent resource constraints, but also ensure robustness against the open and vulnerable wireless medium, as well as emerging threats introduced by AI-driven network intelligence. In particular, such paradigms should be capable of adapting to dynamic environments, preserving the integrity and confidentiality of transmitted information, and supporting task-oriented communication where the relevance of information is prioritized over bit-level accuracy. Addressing these requirements calls for a fundamental shift toward semantic-aware and secure communication frameworks that jointly consider efficiency, intelligence, and security which can be useful in integrated TN–NTN networks. In addition, the distributed and resource-constrained nature of TN–NTN systems requires security mechanisms that are not only effective but also lightweight and adaptive. Traditional security solutions, developed for 5G networks, rely on static configurations and do not account for the dynamic characteristics and limited resources of 6G environments. Therefore, there is a need for adaptive security frameworks capable of dynamically assessing risks and selecting appropriate countermeasures under resource constraints.

0.3 SemCom for TN–NTN Systems

SemCom has recently emerged as a promising paradigm to address the limitations of conventional communication systems. Unlike traditional approaches, which aim to accurately reconstruct transmitted bits at the receiver, SemCom focuses on the meaning and relevance of the transmitted information. The objective is to ensure that the receiver can successfully perform a given task, rather than reconstructing the exact original data.

By prioritizing task-relevant information, SemCom provides reductions in bandwidth consumption, latency, and energy usage. This is particularly beneficial in TN–NTN systems, where communication resources are limited and efficient data transmission is critical. For instance, in cooperative Earth observation scenarios, multiple satellites or aerial nodes may capture visual data from different perspectives. Instead of transmitting full-resolution images or videos, SemCom allows nodes to exchange only the features necessary for tasks such as object detection, classification, or segmentation. This reduces communication overhead while maintaining task performance [2].

The integration of deep learning techniques has further enhanced the capabilities of SemCom. In particular, deep JSCC has been proposed as an effective approach for end-to-end communication system design. JSCC directly maps source data, such as images or videos, into channel symbols using neural networks. In other words, JSCC enables joint optimization of compression and transmission. This approach provides several advantages over conventional separation-based schemes, including improved robustness to channel variations and graceful performance degradation under low SNR conditions [9]. Recent advances in semantic-aware JSCC frameworks incorporate task-specific priors, such as segmentation maps, to guide the encoding and decoding processes. These methods enhance feature discriminability and improve reconstruction quality while maintaining low computational overhead.

While SemCom and JSCC yield significant advantages in terms of efficiency and robustness, their adoption in TN–NTN systems introduces new security challenges. The broadcast nature of wireless communication, particularly in air-to-ground and satellite links, makes transmitted signals inherently accessible to unauthorized users. This exposure is intensified in TN–NTN environments, where transmissions cover wide geographical areas and may be intercepted by adversaries located at different points in the network.

In SemCom systems, security risks are not limited to traditional bit-level interception. Instead, adversaries may exploit the semantic representations of data to infer sensitive information. Moreover, due to the high correlation between transmitted compressed data (latent representation) and original data, an eavesdropper can potentially reconstruct meaningful content from intercepted signals. This type of threat is particularly severe in AI-driven systems, where neural networks can be used to approximate the decoding process.

As discussed in [10] , SemCom systems are vulnerable to a range of attacks, including eavesdropping, inference attacks, and MIEA. In MIEA, an attacker captures the transmitted signal and trains a surrogate model to reconstruct the original data which compromises the privacy of the communication. These attacks highlight the limitations of traditional security mechanisms, which are primarily designed for bit-level protection and do not account for the semantic content of transmitted data.

To address these challenges, several defense strategies have been proposed, including physical-layer security, cryptographic methods, data-driven approaches, and semantic covertness. However, many of these solutions either incur high computational overhead, degrade communication performance , or are time-consuming, making them unsuitable for resource-constrained TN–NTN systems.

0.4 Adaptive Security in 6G TN–NTN Systems

In addition to communication-level vulnerabilities, TN–NTN integration introduces broader cybersecurity challenges due to the scale, heterogeneity, and dynamic nature of 6G systems. Devices operating in aerial and satellite segments often have limited computational capabilities, energy resources, and storage capacity, which restrict the deployment of complex security mechanisms. Furthermore, the diversity of devices and services in TN–NTN systems leads to varying security requirements, making static, one-size-fits-all security solutions of 5G ineffective [11].

To address these challenges, adaptive and risk-based security frameworks have been proposed as a promising approach. These frameworks aim to dynamically assess cybersecurity risks and select appropriate countermeasures based on current system conditions and available resources. By modeling the interactions between assets, threats, and vulnerabilities, risk-based frameworks enable informed decision-making and efficient allocation of security resources.

As demonstrated in [12], such frameworks can be formulated as optimization problems that balance the trade-off between security effectiveness and resource constraints. This approach is particularly relevant for TN–NTN systems, where efficient utilization of limited resources is essential for maintaining both communication performance and security.

0.5 Research Questions

This thesis aims to develop solutions applicable for TN–NTN systems by efficient JSCC, secure transmission mechanisms, and adaptive cybersecurity frameworks. To systematically address these challenges, the following research questions are formulated:

RQ1. How can multimedia data (especially video) be transmitted efficiently and robustly over wireless channels, considering bandwidth limitations, channel noise, and task-oriented requirements?

RQ2. How can semantic information be effectively incorporated into communication frameworks to prioritize task-relevant features and improve transmission efficiency applicable for space–air–ground networks?

RQ3. What are the key security vulnerabilities of SemCom systems in broadcasting nature of TN–NTN environments, and how can they be mitigated against MIEA?

RQ4. How can lightweight and practical protection mechanisms be designed for JSCC-based systems without degrading reconstruction performance or requiring costly retraining?

RQ5. How can cybersecurity risk be modeled and managed dynamically in 6G systems, considering resource constraints and limited budget?

RQ6. How can communication efficiency and security be jointly optimized in TN–NTN systems without introducing excessive computational overhead?

The research questions outlined above are addressed through the contributions of this thesis, which are summarized in the following section.

0.6 Contributions of the Thesis

This thesis investigates efficient and secure communication mechanisms, focusing on SemCom, JSCC, and adaptive security frameworks. The proposed frameworks are applicable to integrated TN–NTN scenarios. The main contributions are as follows:

Chapter 1 provides the necessary background on SemCom, Deep Joint Source-Channel Coding (DJSCC), eavesdropping threats, model inversion attacks, and existing protection mechanisms.

Chapter 2 proposes a lightweight semantic-aware JSCC framework for video transmission, designed for bandwidth-constrained and noisy TN–NTN environments. The framework integrates semantic attention mechanisms and SNR-aware modulation to enhance robustness and transmission efficiency.

Chapter 3 investigates security challenges in SemCom systems, with a focus on eavesdropping and MIEA. A novel lightweight and reversible protection scheme, KAP-DJSCC, is proposed, leveraging key-based transformations to secure latent representations without compromising reconstruction performance. The proposed secure JSCC frameworks can be applied to TN-NTN, NTN-NTN, and TN-TN scenarios.

Chapter 4 introduces a risk-based adaptive security framework for 6G systems, enabling dynamic selection of countermeasures based on real-time risk assessment under resource constraints. The framework is particularly suited for TN–NTN environments.

Finally, the last chapter concludes the thesis and outlines future research directions.

Chapter 1

Background

1.1 Introduction

As the number of users, devices, and communication protocols continues to proliferate rapidly, the demand for efficient information generation and transmission in next-generation networks has significantly increased [13]. Since the inception of digital communication systems, the primary objective has been to ensure the reliable transmission of data over noisy channels. Shannon's theorem has been highly influential in advancing communication systems, providing a theoretical foundation for determining channel capacity. It addresses not only the idealized scenario of noise-free communication but also extends to practical cases where real-world channels are affected by noise, thereby guiding the design and optimization of contemporary communication technologies [14]. Specifically, if the entropy of a source is less than the channel capacity, it is possible to design a coding system that enables transmission with an arbitrarily low probability of error. To address this and mitigate practical challenges such as bandwidth limitations, transmission errors, and fluctuations in channel conditions, advanced source and channel coding schemes have been developed to ensure accurate and reliable data transmission. Nevertheless, the fundamental challenge in Shannon's communication theory is to minimize uncertainty in data transmission while simultaneously maximizing capacity, enhancing reliability, and reducing latency. Addressing these requirements often leads to increased bandwidth consumption and other resource demands, which can constrain the practical implementation of communication services. Consequently, researchers have concentrated on optimizing resource utilization to approach the theoretical limits of communication while balancing efficiency with system constraints. Moreover, according to Shannon's theorem, communication methods primarily focus on ensuring accurate transmission at the bit level, irrespective of the meaning or content of the transmitted data. The emphasis is solely on the reliable delivery of bits, independent of their semantic interpretation [14].

To address these challenges, Shannon's pioneering approach to communication theory introduced a novel conceptualization of communication problems, as detailed in his subsequent work [15]. This expanded framework offered a deeper understanding of data transmission in noisy environments, significantly influencing the evolution of modern communication systems. Therefore, effective communication encompasses more than merely ensuring the accuracy of transmitted symbols; it also involves addressing semantic and pragmatic aspects, including the accuracy of interpreting the transmitted symbols and the extent to which the conveyed information influences the recipient's behavior or actions. This broader perspective underscores

the importance of understanding not only how data is transmitted but also how it is comprehended and applied by the receiver. Considering the semantics of the message, the focus shifts toward the effectiveness of conveying the intended meaning and achieving shared objectives. This suggests that the precise transmission of all data may not always be necessary; instead, transmitting only the information that is semantically meaningful can be sufficient to maintain system performance [16]. This approach emphasizes efficiency by prioritizing the delivery of relevant data, potentially reducing resource consumption while maintaining communication effectiveness. As users generate vast amounts of information, focusing on the meaning of the data becomes increasingly critical, especially given the substantial bandwidth requirements for transmission under low-latency constraints. Indeed, strictly adhering to the coding and transmission methods outlined by Shannon [14] may result in network congestion, potentially exceeding the system's capacity and leading to problematic slowdowns. Therefore, the development of data-aware systems that prioritize extracting and transmitting only task-relevant information—without the need to send and retrieve the entire dataset—will be crucial for future communications [17]. The wide range of applications, including the Internet of Things (IoT) communication, industrial scenarios, and more broadly massive Machine-Type Communications (mMTC), entails diverse requirements and data that are highly dependent on the specific application context. This diversity shifts the emphasis from reconstructing exact messages to reasoning, selecting, deciding, and acting upon specific tasks. This paradigm highlights the importance of meaning in communication and drives the emergence of SemCom [18] and task or goal-oriented communication as effective solutions [19]. Regarded as a significant advancement in bit-wise communication, SemCom has the potential to reduce data transmission load, improve efficiency, optimize resource allocation, enhance reliability, significantly lower latency, and decrease energy consumption [20]. These benefits contribute to enhanced sustainability and increased communication system speed [16], positioning SemCom as an optimal solution for applications requiring extensive data transmission and connectivity, such as object classification for autonomous driving [21] and video streaming [22].

Furthermore, recent advancements in AI have significantly expanded the potential for integrating semantic and task-oriented approaches into the design of communication systems. The incorporation of deep learning, a fundamental branch of AI, into SemCom offers substantial improvements in network efficiency by accurately extracting, interpreting, and optimizing the meaning of source data for transmission[23].

However, although semantic communication bestows advantages such as lower latency, reduced bandwidth consumption, and higher throughput [24], it remains vulnerable due to the inherent openness of wireless broadcast channels and the fragility of AI models [25]. The accessibility of wireless signals to both legitimate and unauthorized receivers raises significant security concerns, such as spoofing [26] and man-in-the-middle attacks [27]. Moreover, AI architectures learn from available examples and transmitted data, potentially exposing sensitive information that can be inferred by adversaries. Additionally, blind spots within AI frameworks can be exploited by attackers through perturbation injection, allowing them to deceive the system [28], [29].

Given these vulnerabilities in wireless networks and the risk of interception by

unauthorized parties, SemCom, despite its advantages, remains susceptible to eavesdropping. Attackers may employ techniques such as wiretapping to capture and intercept the information during transmission, using their knowledge to reconstruct the target sensitive data. The availability of the databases and the fragility of AI models also makes the SemCom model vulnerable to MIEAs in which the eavesdropper can illegally use the encoder and train its network to infer the meaning. Recent research has focused on developing new methodologies or adapting existing bit-wise schemes to the semantic domain to mitigate such attacks. The strategies predominantly explored include Physical Layer Security (PLS) [30], [31], [32], [33], [34], cryptographic mechanisms [35], [36], [37], [38], [39], [40], data-driven methods [41], [42], [43], [44], [45], [46], and covert SemCom [28], [47], [48]. These approaches aim to either increase the secrecy rate [30], [33], [34] or reduce the amount of sensitive information or semantic recovery available to unauthorized receivers [42], [43], [44], [47], [49].

Despite the valuable contributions of existing works addressing various aspects of SemCom, research on privacy preservation, eavesdropping, and MIEA remains in its infancy. Existing review papers and surveys have primarily focused on the development of SemCom theory, network types, architectures, frameworks, system design, algorithms for semantic extraction, transmission methods, and potential applications [8], [16], [17], [20], [50], [51], [52]. In addition, the studies in [21], [22], [32], [50], [53], [54], [55], [56] provide brief discussions of security challenges in SemCom; however, they largely overlook the specific problem of eavesdropping and do not offer a comparative analysis of existing mitigation strategies. Moreover, The work in [57] provides a comprehensive overview of the SemCom lifecycle, addressing its security challenges and defense mechanisms at both the link and network levels. Nonetheless, its discussion of eavesdropping remains general and lacks a detailed review of concrete solutions and countermeasures against semantic eavesdroppers. Similarly, [58] reviews steganographic techniques related to eavesdropping in SemCom; however, it neither compares traditional and semantic eavesdropping nor examines MIEA, and presents only a limited assessment of mitigation approaches within the SemCom framework.

1.2 Background on Semantic Communication and Security Aspects

The primary components of a typical SemCom architecture, inspired by the reference model outlined in [55], are depicted in Figure 1.1. Without loss of generality, the information source (commonly referred to as Alice) transmits information to the destination (referred to as Bob). In a SemCom scenario, Alice and Bob must share the same semantic perception of the message, even if the transmitted and received data are not identical at the bit level. The figure illustrates that the Knowledge Bases (KBs), the semantic layer, and the physical layer are responsible for extracting and transmitting the meaning, ensuring that the intended perception of the message is accurately conveyed. The integration of AI into SemCom significantly enhances the interpretation and transmission of information, marking a major advancement in

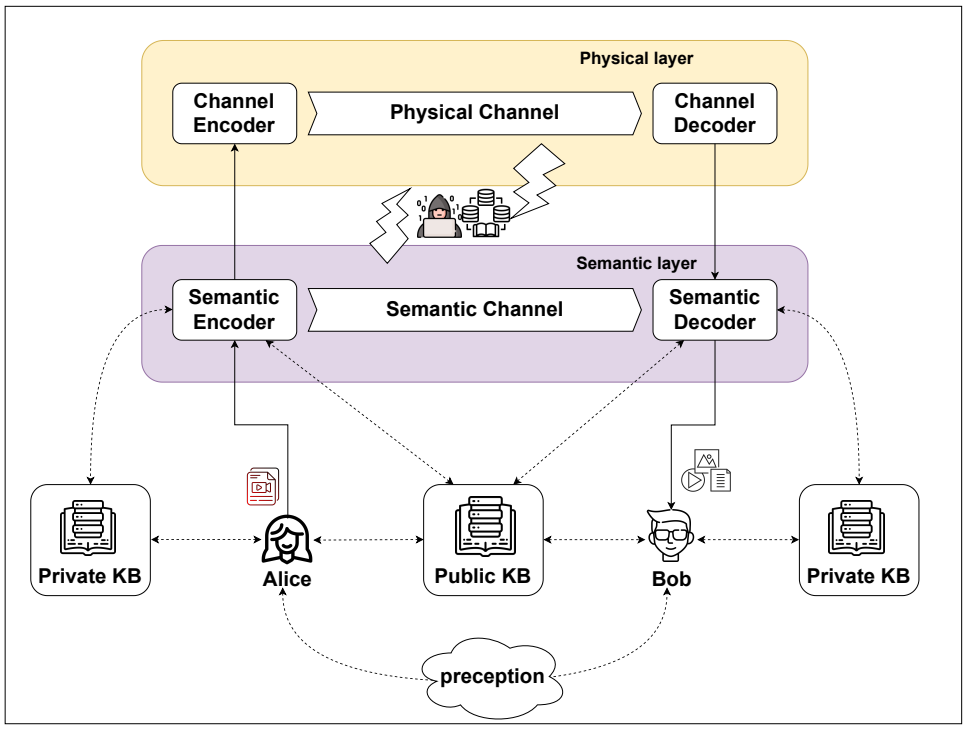

FIGURE 1.1: The SemCom architecture inspired by the reference model in [55].

communication technology. Specifically, useful features of data generated by the information source (e.g., images, video, text, and multi-modal data) are processed in the semantic encoder and reconstructed by the decoder. Both the semantic encoder and decoder operate within the semantic layer, which is responsible for extracting semantic features, transmitting them through the semantic channel, and subsequently reconstructing them. In this layer, AI enhances information recovery and feature extraction by leveraging both public and private KBs. Public KBs provide common knowledge shared among all users, whereas private KBs encapsulate user-specific background information [59]. Overall, KBs play a crucial role in extracting key features by supplying the AI model, hardware, and methodologies to both Alice and Bob. These KBs can be updated as new information becomes available.

Nevertheless, while KBs are essential for the performance of SemCom, they also introduce significant security challenges. The integration of AI for feature extraction creates vulnerabilities, enabling adversaries to forge or poison the model and data by targeting and contaminating the KBs or accessing and using KBs for their illegal purposes. Furthermore, the subsequent wireless transmission phase exposes the system to potential semantic manipulation and unauthorized access to the transmitted information.

In this context, security attackers can target critical system assets, compromising the fundamental security requirements necessary to maintain a secure framework. The primary security requirements of a SemCom framework, as detailed in [55], are

outlined below:

- Data confidentiality: ensures that unauthorized entities cannot access or disclose the data.
- Data integrity: ensures the correctness, consistency, and reliability of data in terms of semantics between Bob and Alice.
- Data availability refers to the accessibility and reliability of data at each stage of SemCom, ensuring that the data remains unaffected by noise or disruptions.
- Data Authenticity: guarantees the trustworthiness of semantic features throughout all stages of SemCom, with verification at the receiver's end to ensure the information has not been falsified.
- Privacy Protection: ensures that Alice's private information is not exposed to other entities or through public KBs.

To explore potential disruptions to these security requirements, researchers have primarily examined four types of attacks in SemCom. First, adversarial attacks, referred to as semantic noise, threaten both the information source and the physical communication channel. Second, poisoning attacks involve adversaries that contaminate the information by injecting malicious data, thereby compromising the integrity of transmitted content. Next, network attacks are characterized by the misuse of transmitted semantic features, which may involve extracting or exploiting information, as well as depleting network devices and resources. Finally, inference attacks leverage transmitted semantic features, combined with shared or publicly available knowledge, to deduce sensitive information about the sender.

Table 1.1 summarizes the main categories of attacks in SemCom [55]. To further emphasize the significance of these threats, the table also highlights the corresponding defense strategies and countermeasures proposed to mitigate these potential vulnerabilities.

TABLE 1.1: Overview of key threats in SemCom and the corresponding defense mechanisms proposed to mitigate them.

Threat	**Defense**
Adversarial attacks (semantic noise)	Adversarial training, model optimization, denoising.
Poisoning attacks	Denoising, model pruning, data preprocessing, detecting training abnormalities.
Network attacks	Cryptography, secure topology, digital signature, traffic monitoring, enhanced planning [55], PLS, semantic covertness, data-driven methods.
Inference attacks	Anonymization, differential privacy, homomorphic encryption [55], cryptography, PLS, semantic covertness, data-driven methods.

Two distinct characteristics in security issues in SemCom differentiate it from bitwise communication. Firstly, the meaning of the captured information is important

to the warden instead of the amount of stolen information. Secondly, the attacker can target both the transmitted sequence and the AI model [53].

The main focus of this chapter is on eavesdropping attacks in SemCom, where the attacker can access the wirelessly broadcasted signal, attempt to reconstruct it, and infer the targeted information. According to [55], eavesdropping can be classified as a type of network attack as well as an inference attack. In addition, MIEA, as a special case of eavesdropping attacks in the semantic domain, is analyzed, where an eavesdropper captures the signal and attempts to reconstruct it by training an inverse Deep Neural Network (DNN) network.

As shown in Table 1.2, a wide range of review studies have explored various aspects of semantic communication, including its security vulnerabilities, attack models, mitigation techniques, and related security requirements.
It highlights the main objectives of each review and identifies whether security concerns are analyzed in depth, only briefly mentioned, or suggested as future research directions. The table also indicates whether eavesdropping is explicitly discussed, how it is characterized, and whether corresponding mitigation strategies are proposed. As shown in the fifth column, eavesdropping is rarely treated as an independent topic; instead, it is typically grouped with other security issues such as adversarial, jamming, or man-in-the-middle attacks, and addressed through shared countermeasures. With the exception of [8], which briefly contrasts eavesdropping in semantic and conventional communication, prior works do not provide a focused comparison or taxonomy of existing eavesdropping techniques in SemCom.

However, while an extensive review of fundamental semantic definitions, theories, algorithms, applications, and challenges is presented, the discussion on wiretapping remains limited, despite the potential for adversaries to access and reconstruct targeted information using their KB. To the best of our knowledge, no study has performed a comparative analysis of existing works in the context of eavesdropping and MIEA. Thus, based on the referenced surveys, it is evident that no existing research specifically investigates eavesdropping and the resulting privacy leakage in SemCom, nor does it provide a comparison of the state-of-the-art methods in this area against a new eavesdropping attack designed for SemCom. This document not only categorizes the existing approaches for semantically limiting the warden's capabilities, but also distills the key lessons learned from prior approaches, emphasizing their strengths, limitations, and the critical need for robust defense mechanisms against semantic eavesdropping.

At the time of this writing, and to the best of the authors' knowledge, there is a significant lack of review papers specifically addressing privacy leakage, eavesdropping, and wiretap channel mitigation methods in SemCom. A comprehensive investigation and comparison of existing approaches in these areas remain unexplored. To bridge this gap, the main contributions of this chapter are summarized as follows:

- This chapter reviews the fundamentals of SemCom, primarily focusing on aspects related to common attacks and associated security requirements. In particular, the study investigates existing countermeasures for eavesdropping in traditional communication and explores the potential for adapting these solutions to semantically-based communication.

TABLE 1.2: Comparative overview of major review papers on SemCom: A — Security coverage, B — Eavesdropping focus, C — Comparison between semantic and traditional eavesdropping.

Ref.	**Year**	**Main Focus**	**A**	**B**	**C**	**Technical Comparison**
[16]	2022	Information-theoretic foundations, layered architecture, performance metrics, semantic noise, challenges	No	No	No	Bilingual Evaluation Understudy (BLEU) vs. JSCC in semantic coding
[17]	2023	History and theory of SemCom, JSCC, age of information, and timing aspects	No	No	No	No
[20]	2022	Fundamentals, semantic extraction, transmission, metrics, applications, challenges	Yes	No	No	No
[21]	2024	Task-oriented communication, Integrated Sensing and Communication (ISAC), multi-task learning, vulnerabilities	Yes	No	No	Task accuracy under adversarial attacks
[50]	2021	Semantic-aware networking, federated edge intelligence for SemCom	No	No	No	Comparison across edge server scenarios
[51]	2021	Categorization of SemCom domains, knowledge graphs	No	No	No	No
[52]	2025	Fundamentals, semantic entropy, algorithms for SemCom security, resource allocation, quantum aspects	Yes	No	No	No
[8]	2024	Three-tier SemCom architecture, taxonomy of security challenges and countermeasures	Yes	No	No	No
[53]	2024	Semantic extraction (autoencoder, IB, KG), attack strategies, and mitigation methods	Yes	No	No	No
[54]	2022	Overview on semantics-aware communication, joint coding and RL-based metrics	Yes	No	No	RL-based evaluation of SemCom
[55]	2023	Threats and defenses in SemCom across technical, semantic, and effectiveness levels	Yes	No	No	Privacy protection vs. FedAvg under MIA
[22]	2021	SemCom in 6G, use cases, KPIs, energy-delay trade-offs	No	No	No	DeepSC vs. traditional coding
[32]	2023	Secure SemCom for metaverse, privacy challenges	Yes	No	No	Accuracy of proposed vs. FedAvg
[56]	2023	Security in semantic IoT, novel KPIs	Yes	Yes	No	Semantic similarity in targeted poisoning attack but not eavesdropping, evaluation of the proposed KPIs in outage probabilities
[57]	2025	Link- and network-level perspectives, security in model training and transfer	Yes	Yes	No	No
[58]	2025	General image encryption and steganography techniques, their potential in SemCom, relevant methods, and future directions	Yes	Yes	No	Coverless steganography scheme
Our work	**Now**	**Comprehensive review of SemCom vulnerabilities with focus on eavesdropping, comparison with traditional wiretapping, analysis of mitigation methods and introduction of MIEA**	**Yes**	**Yes**	**Yes**	**Analysis of MIEA scenarios and mitigation techniques**

- Additionally, this work provides an overview of existing strategies specifically designed to defend against potential eavesdroppers in SemCom. To offer further insights, the proposed study categorizes the mitigation schemes and the metrics used for security measurement.
- A comparative analysis is conducted in a general AI-aided SemCom scenario under various wiretap channel SNRs under MIEA to identify potential gaps for future research. This analysis provides a qualitative evaluation that highlights the strengths and weaknesses of the observed strategies, offering valuable insights into the effectiveness of these approaches in mitigating privacy leakage caused by eavesdroppers and similarity measurement in the semantic domain.

The remainder of this chapter is organized as follows: Section 1.3 highlights the key differences in the impact of eavesdropping attacks between traditional communication systems and SemCom. Section 1.4 reviews and categorizes the state-of-the-art defense mechanisms proposed to counter eavesdropping attacks in SemCom. Finally, Section 1.5 concludes the chapter with closing remarks and key insights.

1.3 Eavesdropping Issues in Traditional and Semantic Communication

The broadcasting nature of wireless communication presents an opportunity for unauthorized users to intercept transmitted signals. In this context, an eavesdropper is defined as an illegitimate user positioned between the legitimate signal transmitter and receiver, seeking to overhear the signal and wiretap the information. Due to the intrinsic differences between bit-wise communication and SemCom, the eavesdropper's effectiveness must be evaluated according to the distinct characteristics of each communication paradigm. Accordingly, the methods employed by an eavesdropper to intercept and reconstruct transmitted information, as well as the countermeasures devised to mitigate such attacks, vary significantly between these frameworks. In conventional communication systems, an eavesdropper is deemed successful if it captures the transmitted signal with sufficient quality to accurately reconstruct the source message [60]. If an eavesdropper is positioned within the transmission range and receives the signal with adequate quality, it can intercept and retrieve the transmitted bits. To defend networks against such interception, various mitigation techniques have been developed specifically for traditional bit-wise communication systems. In one approach, the data transmitter can obscure the signal by introducing artificial noise within covert communication, proactively safeguarding the transmission and disrupting unauthorized receivers. Covert communication conceals transmitted data within environmental coefficients, thereby masking the transmission from eavesdroppers. However, this approach is constrained by variability in channel coefficients, which may affect reliability [61]. In contrast, PLS methods primarily aim to lower the Signal-to-Interference-plus-Noise Ratio (SINR) at the eavesdropper's end, degrading the intercepted signal's quality and preventing accurate reconstruction. These methods require minimal computational resources for authorized users but depend on additional information, such as the eavesdropper's channel state, which may be difficult to obtain and could limit practical applicability. Cryptographic schemes obfuscate the

signal by converting it into an encrypted format using private and public keys or hash functions. Although these schemes provide a proactive defense by rendering the data unreadable to unauthorized users, they require significant computational resources to maintain data security [62]. Lastly, spread spectrum communication techniques enhance security by transmitting the signal across a bandwidth larger than necessary, reducing susceptibility to narrowband interference and enhancing resistance to interception [63].

However, these techniques focus solely on bit-wise data transmission and identical bit reconstruction at the receiver, with the primary goal of degrading or disrupting the eavesdropper's ability to accurately reconstruct the transmitted data.

With the change to a semantic communication paradigm, eavesdropping becomes more challenging, since an attacker must first intercept the transmitted signal and subsequently decode its semantic content to access specific sensitive information [56].In this context, the eavesdropper may possess a different KB or a mismatched trained decoder, which could result in errors during data reconstruction or misalignment with the intended communication task.

In SemCom, an attacker must not only intercept the transmission but also possess the capability to infer sensitive information, which may require understanding complex AI auto-encoder models or having prior familiarity with the KBs used in semantic processing. This originates from including AI and KBs in SemCom which results in a modified system design, security threats, and countermeasures compared to the traditional communication. To address security and privacy challenges in SemCom, a hybrid approach that integrates traditional and AI-based techniques is employed to ensure secure information transmission within the semantic network. Innovative methods, inspired by conventional communication paradigms and designed to safeguard semantic data, will be detailed in the following sections.

1.4 Current trends in SemCom eavesdropping defense methods

As discussed in the preceding section, the solutions addressing eavesdropping attacks in traditional communication systems cannot be directly extended to SemCom. These solutions require careful revision and additional considerations to effectively address the unique challenges posed by SemCom. In response to this critical issue, several metrics and methodologies have recently been proposed in the current state of the art.

With the potential integration of SemCom into mMTC and industrial IoT applications, a key challenge lies in redefining existing frameworks and evaluation metrics to reflect task-specific objectives and semantics. The choice of metrics is particularly crucial, as they are inherently linked to the content and intent of semantic IoT communications. On one hand, current evaluation metrics are largely inherited from traditional communication systems. For instance, in image transmission, metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM) , MS-SSIM, Mean Squared Error (MSE), classification accuracy, and task success rate are commonly used to assess pixel-level reconstruction quality or task-specific performance. In addition, traditional image processing metrics such as Local Feature Based Visual Security (LFBVS) [64] and Correlation (COR) [65] have been

employed to evaluate contextual aspects of visual security [42]. While these metrics provide insights into perceptual and structural fidelity, they were not designed to capture the semantic relevance or task-specific utility of the transmitted content. Meanwhile, the Learned Perceptual Image Patch Similarity (LPIPS) metric [47] leverages deep learning models to offer a more semantically aligned assessment of similarity between images. Furthermore, the Matthews Correlation Coefficient (MCC) [66] introduces a task-oriented approach for evaluating attribute prediction performance, particularly under conditions of imbalanced label distributions. The information leakage metric proposed in [67], which is based on mutual information, quantifies the amount of sensitive information exposed to potential eavesdroppers in transmitted images. This highlights a critical privacy concern in semantic communication systems, where the extraction and transmission of high-level semantic features may inadvertently reveal private or task-specific information, necessitating the development of privacy-preserving mechanisms tailored to semantic content. Similarly, the work in [38] introduces information leakage as an evaluation metric in the context of wireless text transmission. To further elaborate on metrics for text-based communication, semantic similarity is commonly assessed using metrics such as BLEU, S-BLEU, and sentence-level similarity scores. Among these, BLEU has been particularly influential in transitioning traditional surface-level evaluation methods toward more semantically oriented assessments of sentence interpretation. The secrecy rates proposed in [30], [33], [34], [68] leverage BLEU to incorporate semantic meaning, thereby adapting the traditional secrecy rate formulation to better suit application-specific requirements. Without loss of generality, Table 1.3 summarizes the most widely used metrics for evaluating semantic information recovery and quantifying information leakage to potential eavesdroppers.

TABLE 1.3: Summary of the metrics employed in the literature.

Data	**Metric**	**Reference**
Img	PSNR	[35], [37], [41], [47], [69] [28], [42], [44], [66], [70], [71]
	SSIM, MS-SSIM	[28], [35], [37], [42], [44], [45], [46], [47], [71]
	MCC	[66]
	Image visual quality	[31], [37], [41], [44], [66], [70] [42], [43], [46], [47], [67], [69]
	MSE	[36], [72]
	Accuracy and task success rate	[32], [41], [43], [45], [67] [31], [44], [70], [71], [72], [73]
	LPIPS	[47]
	COR	[42]
	LFBVS	[42]
	Information leakage	[67]
Txt	BLEU, sentence similarity	[33], [38], [40], [74], [75]
	Achievable network throughput	[34]
	S-BLEU	[74]
	Information leakage	[38]
	Secrecy rates	[30], [33], [34], [68]

Beyond evaluation metrics, this section categorizes existing approaches for securing SemCom in wiretap channels into four main groups: PLS, encryption and secret key management, data-driven techniques, and methods aimed at disguising semantic information and ensuring covertness. Furthermore, innovative methods that deviate from these main categories or can be generalized to SemCom are explained in detail in Section 1.4.5.

1.4.1 PLS methods

Approaches for safeguarding SemCom that exploit the dynamics and randomness inherent in wireless channel characteristics are broadly categorized under PLS methods [74]. Unlike conventional communication, PLS methods in SemCom primarily involve manipulations and optimizations of the physical components to evaluate the task performance indicators and obstruct Eve in semantics and interpretations. These include techniques such as Artificial Noise (AN) generation and friendly jamming [34], [48], [75], [76], beamforming [30], [31], [33], and the use of Reconfigurable Intelligent Surface (RIS) [30], [31], [36], [68] along with optimizing transmitting or AN power [33], [34], [48], [68], [75] which are employed to evaluate task performance indicators and impede eavesdroppers at the semantic layer.

Figure 1.2 presents a general framework for the aforementioned methods, which are integrated following semantic encoding and jointly optimized. By leveraging these methods, although Eve may receive the signal, it remains highly degraded in her direction or lacks sufficient quality for accurate task interpretation. This part of the chapter provides an overview of PLS methods defending from eavesdropping attacks in SemCom, summarized in Table 1.4.

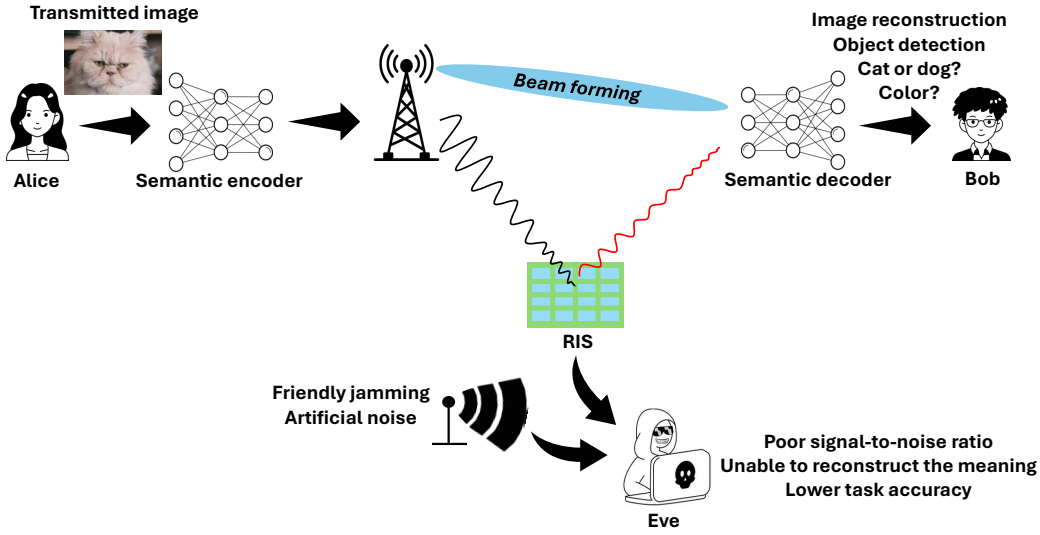

FIGURE 1.2: PLS methods general framework.

Similar to conventional communication, SemCom can be secured using AN and its combinations with other techniques. Mu et al. [34] explored the potential of PLS by proposing a semantic-assisted approach. In their study, Eve intercepts and decodes

TABLE 1.4: PLS methods taxonomy.

Ref.	Data Type	AN	Power Allocation	RIS	Beamform.
[33]	Txt	✓	✓		✓
[30]	Txt			✓	✓
[34]	Txt	✓	✓		
[31]	Img	✓	✓	✓	✓
[36]	Txt			✓	
[68]	Txt		✓	✓	
[76]	Txt	✓	✓		
[75]	Txt	✓	✓		
[48]	Img	✓			

signals using a bit-oriented approach, whereas Alice and Bob employ systems capable of both semantic and bit-oriented processing. By transmitting a superimposed signal comprising normalized semantic-encoded symbols and normalized conventional bit-oriented source–channel encoded symbols over a block fading channel, the semantic component effectively functions as AN, thereby enhancing confidentiality. The achievable secrecy communication rate for a given fading state is determined using DeepSC [77] for semantic text transmission. The ergodic secrecy rate is maximized by jointly optimizing the transmit power, power allocation factor, and successive interference cancellation decoding order for each fading state, formulated as a non-convex optimization problem. Building on the use of AN, the authors of [76] proposed an AN-aided covert and reliable SemCom scheme, leveraging a dual-antenna configuration at Bob's side to enable full-duplex reception. One antenna is dedicated to receiving the signal, while the other generates AN to mitigate continuous eavesdropping. The framework optimizes the average covert semantic throughput—a newly introduced metric—across all time slots using reinforcement learning for semantic text transmission, as proposed in [77].

AN, acting as a friendly jammer, has been utilized in [75] to obscure transmission. The proposed approach employs a soft actor-critic algorithm to maximize the BLEU score while simultaneously regulating both transmission and jamming power. Additionally, it ensures that the probability of missed detection at Eve exceeds a predefined threshold, thereby enhancing covert communication security. A jammer-aided SemCom method for multi-user communication networks within a multi-agent reinforcement learning framework has been proposed in [48]. The multi-agent gradient algorithm enhances both performance and covertness by collaboratively identifying vulnerable devices, optimizing transmit power, and refining semantic information transmission. Additionally, the distributed algorithm establishes a power control policy, trained by both the devices and the jammer, to optimize SemCom performance.

Besides using AN to conceal the communication, integrating it with beamforming techniques and jointly optimizing them is an effective adopted approach for mitigating eavesdropping attacks. The work in [33], these solutions were applied to ISAC with multiple users—referred to as Integrated Sensing and Semantic Communication (ISSC)—by optimizing beamforming and leveraging the sensing signal as AN. In

fact, the sum of the semantic secrecy rate has been maximized by iteratively optimizing beamforming for sensing and communication, power allocation, and compression ratio, while considering the BLEU score as the performance metric.

Apart from AN, techniques leveraging RIS have also demonstrated effectiveness in this research field. In [30], the mitigation of information leakage in metamaterial-based task-oriented communication and computation has been investigated. Reconfigurable Intelligent Computational Surface (RICS), a technology comprising a reconfigurable beamforming layer, an intelligent computation layer, and a control layer, is employed to destructively combine the non-reflected signal with the refracted signal in the eavesdropper's direction, effectively neutralizing potential leakage. The authors anticipate that metasurfaces will integrate task-oriented communication and computation in the future. Additionally, dynamically adjusting signal reflection through RICS-Design-B further degrades the signal quality at the eavesdropper.

Another type of RIS, known as Simultaneous Transmitting and Reflecting Reconfigurable Intelligent Surface (STAR-RIS), is utilized in [31] to enhance signal transmission for SemCom between Bob and Alice while simultaneously generating interference for the eavesdropper. The study considers a scenario where Alice is positioned in the transmission region of the STAR-RIS, while Eve is located in the reflection region. This is achieved by formulating an optimization problem that jointly optimizes Bob's beamforming vector and the transmission-coefficient vector of the STAR-RIS, while designing the reflection-coefficient vector to induce task-level and SNR-level disturbances for Eve. The generated task-level disturbance surpasses other benchmarks in preserving SemCom privacy, leading to significantly lower task success rates for Eve.

In [36], a novel PLS key generation scheme leveraging RIS is proposed, wherein the RIS elements are dynamically tuned based on the randomness of semantic drifts in text transmission. The physical key generation framework consists of four phases: channel detection, quantization, message coordination, and privacy amplification.

Moreover, [78] proposes an alternative physical-layer key generation method that exploits the randomness of BLEU scores and strengthens security by implementing subcarrier obfuscation in OFDM through dummy data injection. Building on these approaches, [68] explores a secure semantic communication framework for UAVs, leveraging multiple RISs to enhance security. The study focuses on maximizing the Security Semantic Transmission Rate (SSR) by jointly optimizing the RIS reflection coefficients, UAV trajectory, transmission power, and the number of transmitted symbols for semantic text broadcasting.

The work in [74] utilizes secure channel capacity concepts to develop a PLS-empowered loss function that balances security and reliability. This approach, independent of key generation, enhances robustness against channel fading by employing a two-phase DNN training process. Furthermore, this study introduces a novel metric, Secure BLEU, to evaluate the security of semantic communication. Existing approaches, such as those in [76] and [68], are limited by the unrealistic assumption that the eavesdropper's location is known. Unlike traditional encryption techniques, PLS methods ensure security without relying on key secrecy and remain robust under channel fading and variations [74].

1.4.2 Encryption and Secret Key Management

DNNs can learn to utilize secret keys to protect the privacy of SemCom networks. Specifically, they have the potential to autonomously learn encryption techniques and safeguard information, laying the foundation for further advancements in secure communication using AI [79]. This section explores existing works that employ encryption techniques [35], [39], [40], [71], shuffling, permutation, and substitution [37], [38], as well as quantum cryptographic security [80] and secret key management in Federated Learning (FL)-aided SemCom [81].

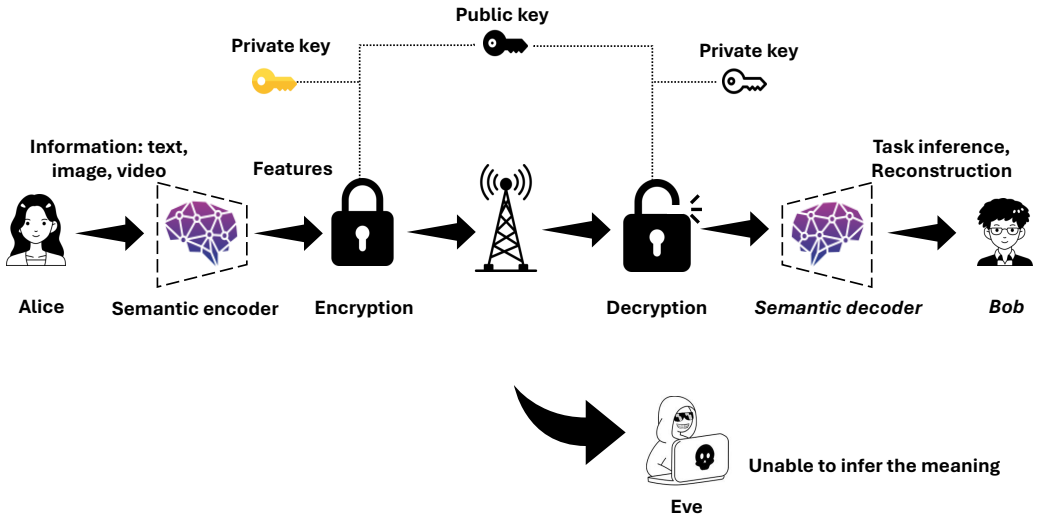

FIGURE 1.3: Encryption and key management methods general framework.

TABLE 1.5: Encryption and key management techniques taxonomy.

Ref.	Data Type	Colluding Eve	Key Management	Encryption	Shuffling and Permutation	Quantum
[37]	Img				✓	
[40]	Txt			✓		
[39]	Img			✓		
[35]	Img			✓		
[80]	NA		✓			✓
[38]	Txt				✓	
[81]	NA	✓	✓			
[71]	Img		✓	✓		

Figure 1.3 illustrates the general encryption framework used for SemCom. By applying cryptographic techniques, Eve is prevented from accurately interpreting the semantic information. Additionally, Table 1.5 provides an overview of the key methods and assumptions explored in what follows.

Cryptographic techniques offer the advantage of transforming data into an unreadable format, ensuring that unauthorized users cannot access its contents, regardless of channel conditions. Regarding the use of cryptography in SemCom, Luo et al. [40] proposed a neural network-based symmetric encryption privacy-preserving technique for text communication. Their approach integrated symmetric encryption into the adversarial training process to ensure both the universality and confidentiality of semantic data. To maintain the universality of the framework, the cross-entropy function of unencrypted data is optimized, particularly in broadcast scenarios. In encrypted SemCom, however, the optimization objective shifts toward maximizing the discrepancy between the cross-entropy loss experienced by Alice and Bob and that of Eve, thereby enhancing confidentiality.

Conversely, safeguarding semantic information can be achieved with a shorter key length. In [39], the authors optimized the trade-off between the coding rate, sensitive information leakage, secret key rate, and reconstruction distortion, ensuring both efficiency and security in semantic communication. According to Shannon's definition of *perfect secrecy*, Bob must recover the message losslessly, while Eve gains no information. Differently, [39] shows that even a small-sized secret key can effectively protect only the semantic content, despite the observable source data being exposed to an eavesdropper. This highlights a shift from traditional encryption towards a more efficient and targeted security approach in semantic communication.

Moreover, [71] proposes a lightweight and secure SemCom framework for image classification task, aiming to prevent eavesdropping and adversarial attacks. This approach integrates established cryptographic methods, including AES combined with RSA, to effectively counteract wardens and enhance communication security.

Following the same context, the authors of [35] design a novel encryption-based wireless image transmission scheme that operates independently of any knowledge of Eve's channel information in a JSCC scenario. This framework integrates encryption using a public-key cryptosystem based on the Learning With Errors (LWE) problem, ensuring robust security against eavesdroppers without relying on Eve's channel state information. In fact, an encoder function first maps input images into a latent vector, which is then quantized. The quantized representation is subsequently treated as plaintext for encryption, ensuring secure transmission while preserving the integrity of the semantic information.

The security in DNNs can be enhanced by applying shuffling or permutation to the data, effectively obfuscating the information without introducing additional computational overhead for learning or cryptographic processing. In this context, [37] employs a model-based JSCC approach where permutation is applied to the encoded symbols before transmission. This method enhances security by ensuring that only authorized receivers with the correct permutation key can successfully reconstruct the transmitted message, thereby mitigating potential eavesdropping threats. Further on this topic, [42] proposes a novel secure JSCC framework for image transmission that incorporates block-wise shuffling. This technique preserves local spatial redundancies essential for Convolutional Neural Networks (CNNs) while still obfuscating the overall structure of the image, making it more resilient against unauthorized reconstruction by eavesdroppers. Hence, the authors propose an alternative privacy-preserving method for SemCom, leveraging random permutation and substitution of transmitted features to enhance security and mitigate unauthorized access.

Beyond conventional encryption techniques, quantum cryptographic security schemes are expected to play a crucial role in safeguarding future communication networks. Quantum computers pose a potential threat to traditional cryptographic methods used for privacy protection. In [80], quantum key distribution is introduced as a groundbreaking approach for secure secret key distribution, ensuring confidential data transmission, particularly in scenarios vulnerable to eavesdropping. The principles of quantum mechanics are employed to distribute cryptographic keys, which are then used for semantic data encryption. To achieve this, a hierarchical architecture for a key distribution-secured semantic communication system has been designed, integrating resource allocation and routing decisions to enhance security and efficiency.

Finally, FL offers a promising collaborative framework for training joint learning models at the network edge, facilitating the advancement of connected intelligence. However, FL-based SemCom systems remain vulnerable to privacy leakage due to potential eavesdropping threats. A reliable FL system must address two critical concerns: safeguarding user privacy throughout the model training process and ensuring the model's robustness against tampering and adversarial attacks [82]. This necessitates the integration of an additional privacy-preserving mechanism in FL-based SemCom. Given that cryptographic methods are often impractical due to limited computational and communication resources, the authors in [81] propose a Certificateless Authentication-based Trustworthy Federated Learning (CATFL) framework that effectively mitigates poisoning attacks while simultaneously preventing user privacy leakage. To achieve this, the key generation center independently generates two distinct keys for each component within the CATFL framework. Therefore, even if attackers collude with the key generation center or a single server, they cannot obtain the complete secret key required for impersonation.

1.4.3 Data-driven methods

JSCC in SemCom is typically enhanced by a data-driven approach that leverages optimized encoding and decoding functions to exploit the inherent correlation within the source signal for efficient compression [42]. However, many JSCC schemes primarily emphasize improving image or video reconstruction quality, while neglecting their vulnerability to privacy leakage from eavesdropping attacks. This vulnerability arises due to the inherent correlations between the primary data and the channel input, making it susceptible to unauthorized interception [70]. To address this challenge, privacy-aware JSCC has been proposed to optimize the trade-off between image retrieval quality and privacy leakage. In other words, these methods primarily optimize the autoencoder to preserve privacy and adjust a loss function by combining Information Bottleneck (IB) [41], [43], [44], [45], [67], [70] and reconstruction error measurement losses like MSE [41], [42], [43], [44], [45], [46], [67]. These methods consider scenarios in which one or multiple [45] eavesdroppers intercept the noisy wireless signal through their respective channels, attempting to infer sensitive information from the transmitted data.

Figure 1.4 provides an overview of the primary framework in this category, where protection relies on the design of the training process and loss functions. During training, sensitive information is disentangled, and the network is optimized using

loss functions tailored to the specific application and problem requirements. Furthermore, Table 1.6 outlines the key parameters regulated in training the network to safeguard sensitive information. Additionally, it highlights the methods that incorporate multiple eavesdroppers within the loss function to enhance privacy protection.

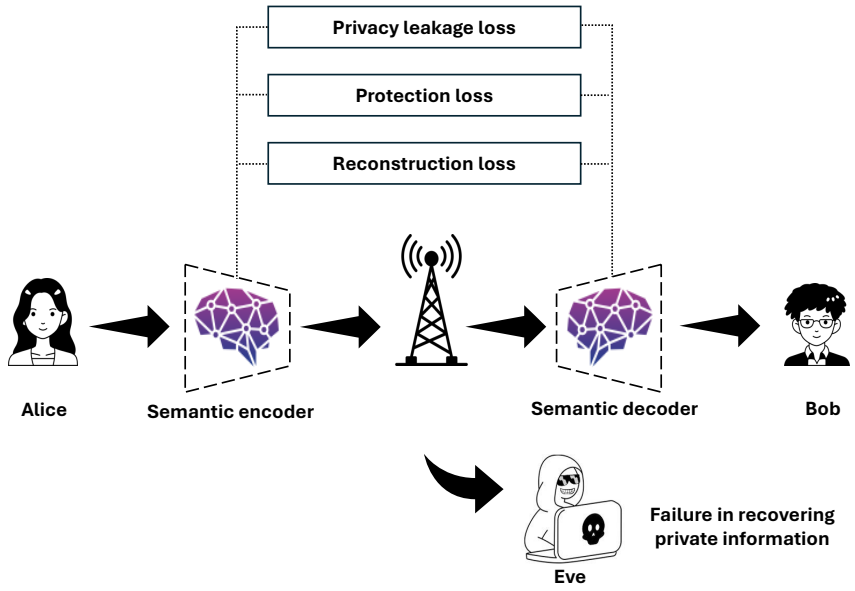

FIGURE 1.4: Data-driven protection training framework and loss functions.

TABLE 1.6: Data-driven methods taxonomy.

Ref.	Data Type	Multiple Eve	IB	Error Measurement
[67]	Img		✓	✓
[45]	Img	✓	✓	✓
[43]	Img		✓	✓
[41]	Img		✓	✓
[44]	Img		✓	✓
[46]	Img			✓
[42]	Img			✓
[70]	Img		✓	

As a foundational study, [41] introduces a privacy-distortion trade-off framework that optimizes the PSNR of the transmitted image as a measure of distortion while simultaneously ensuring that the leakage of sensitive information to Eve remains below a predefined threshold.

In [45], a DNN-based framework is proposed to simultaneously minimize distortion at Bob and mitigate information leakage to multiple collaborating and non-collaborating eavesdroppers. Specifically, a secrecy funnel framework is introduced to safeguard private data while preserving the perceptual quality of transmitted images over AWGN and complex-valued fading channels.

Conversely, [67] introduces an end-to-end learning strategy that simultaneously minimizes distortion and sensitive information leakage against a single eavesdropper

while maximizing channel capacity. Given that true distributions are typically inaccessible, a data-driven approach employing variational autoencoder-based JSCC has been proposed to effectively manage the privacy-utility trade-off. Similarly, another JSCC autoencoder employing CNNs, as proposed in [46], balances efficiency and privacy by utilizing residual blocks cascaded with convolutional layers for semantic feature extraction and transmission. The framework evaluates reconstruction quality and semantic leakage in Single-Input Single-Output (SISO) and Multiple-Input Single-Output (MISO) systems, assessing both visual quality and structural similarity. This approach has inspired further research aimed at optimizing data manipulation techniques to ensure effective transmission for Alice while making reconstruction significantly more challenging for Eve. However, [46] and the aforementioned studies rely on the assumption of precise knowledge of Eve's channel conditions, such as SNR, or operate under the premise that the warden's channel quality is inherently lower. To address the impracticality of these assumptions in real-world scenarios, it is crucial to develop methods that either operate independently of channel state information or dynamically adapt to varying channel conditions. Building on this concept, the authors in [43] introduced a novel privacy-aware JSCC framework that eliminates the dependency on Eve's location or channel state information. Instead, their approach employs a more practical objective function that does not explicitly account for the amount of information received by Eve. In other words, the framework jointly minimizes transmission distortion and the mutual information between private and public disentangled subcodewords, while simultaneously maximizing the similarity between the extracted private information of the image and the generated private subcodeword. Thus, in [43], only the disentangled public information, derived through information bottleneck theory, is transmitted, while the reconstruction distortion at Bob is minimized. This approach ensures that the warden cannot recover the sensitive information, while Alice operates without requiring any assumptions about Eve's presence or channel parameters.

Similarly, the concept of disentangled IB has been employed in [70]. Specifically, the proposed IB objective effectively regulates the mutual information between subcodewords and the encoded information. Minimizing the mutual information between the private and public subcodewords enables JSCC to enhance privacy preservation while simultaneously maintaining the quality of service.

In [44], the authors leveraged IB theory to identify and extract task-relevant features while simultaneously employing adversarial learning to mislead the eavesdropper. Specifically, they modeled the eavesdropper as a DNN-based black-box attacker. A notable aspect of this work is the integration of two key components: the IB loss, which ensures the extraction of task-relevant features while minimizing irrelevant information, and the mean squared error from adversarial reconstruction, which actively disrupts the eavesdropper's ability to accurately recover private data. Similarly, the authors in [66] proposed a data-driven Deep Learning (DL) model that balances the privacy-utility trade-off to enhance system security. This model is designed to encode only the task-relevant information while simultaneously simulating the behavior of a black-box adversary, thereby improving robustness against eavesdropping attacks. In another approach, [42] introduces an inherently secure joint protection and source-channel coding scheme that integrates protection, deprotection, feature extraction,

and JSCC units. They redefine the loss function based on a feature extraction process, aiming to minimize the reconstruction loss while simultaneously maximizing the feature loss between the original image, the protected image, and the decoded image. Several evaluation metrics for visual security assessment [65], [83] and image reconstruction quality indicate that the method proposed in [42] achieves superior reconstruction quality while effectively preserving the visual details of the plain image.

In [84], a KB-based framework is introduced to enhance privacy protection against eavesdroppers. This framework integrates sensitive entity recognition, attention mechanisms, entity alignment, and channel coding modules for secure text transmission. These components collectively constitute a knowledge-assisted transceiver that effectively manages both private and public knowledge, limiting an adversary's ability to decipher sensitive information.

1.4.4 Semantic covertness

This section reviews the state-of-the-art approaches for concealing transmitted data or safeguarding sensitive information. The primary distinction between physical-layer covert communication and semantic covertness lies in their underlying approaches: the former leverages antennas and jamming techniques to conceal the existence of the transmission, while the latter aims to prevent Eve from understanding the transmitted content or to deliberately mislead the warden. In this context, methods such as generating semantic perturbations [28], [72], optimizing covert loss [85], and employing steganographic techniques [47] have demonstrated effectiveness in enhancing semantic covertness and mitigating the risk of information leakage. Figure 1.5 provides a comprehensive summary of the methodologies employed to achieve semantic covertness. In these approaches, Eve may either fail to detect the transmitted stream or be unable to interpret the captured information. Additionally, Table 1.7 presents an overview of the key methodologies utilized in this category, highlighting their respective strategies for concealing semantic information.

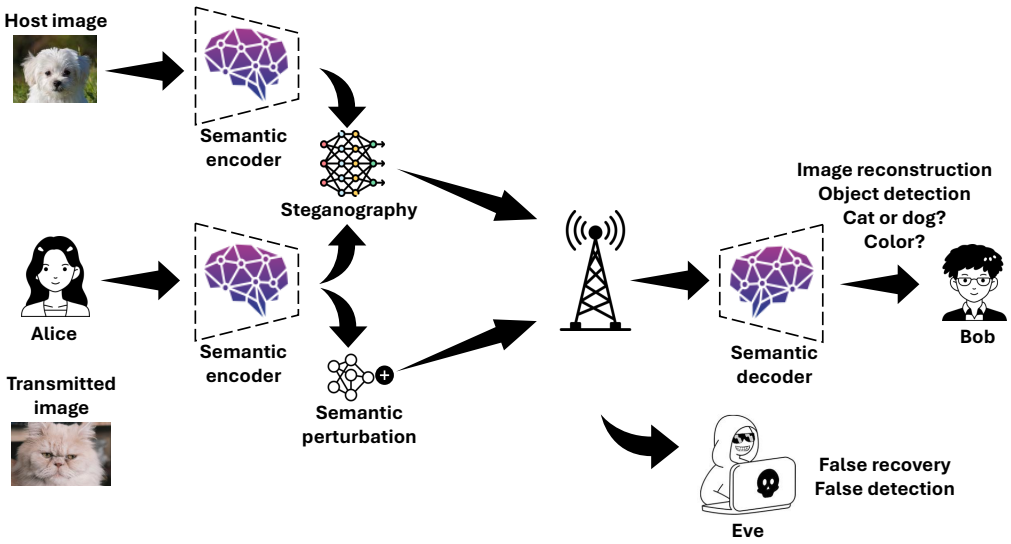

FIGURE 1.5: Semantic covertness methods general framework.

TABLE 1.7: Summary of main features in semantic covertness for SemCom.

Ref.	Data Type	Covert Loss	Steganography	Perturb. generation
[28]	Img			✓
[47]	Img		✓	
[85]	Img	✓		
[72]	Img			✓

In [28], the SemProtector framework is introduced to enhance security in real-world online SemCom. The SemPrIv module in SemProtector is responsible for generating minimal perturbations aimed at misleading potential eavesdroppers. An Adversarial Transformation Network (ATN) is employed to defend against Attribute Inference Attack (AIA), where the attacker is modeled as a multilayer perceptron. The semantic representations are processed by the ATN to generate perturbations that maximize the cross-entropy loss of the eavesdropper, thereby mitigating privacy leakage. Experimental evaluations indicate that while SemProtector effectively enhances security, it slightly degrades performance metrics such as SSIM and PSNR, particularly when multiple protection modules are utilized. Consequently, a trade-off must be made based on user requirements and security priorities.

Building on this concept, [72] introduces pluggable adversarial modules that generate low-power perturbations to mislead Eve in image transmission. This approach optimizes a weighted sum of three key factors: privacy leakage to Eve, mean squared error (MSE) for reconstruction quality, and the power of the adversarial attack. By incorporating these modules, the framework effectively enhances security while maintaining an optimal balance between protection and image quality.

Drawing inspiration from covert communication, [47] proposes a novel approach for secure SemCom in private image transmission through steganography. Unlike cryptographic techniques, image steganography conceals a private image within a non-sensitive host image, producing a container image that visually resembles the host while embedding the private content. If an eavesdropper intercepts the transmitted stream, they can only access the semantic information of the host image, leaving the concealed data protected. The designed architecture leverages an invertible neural network-based signal steganography approach at the channel input signal, effectively mitigating model inversion attacks and enhancing security in SemCom systems.

In [85], the authors explored the fundamental limits of semantic communication under the constraint that transmissions remain covert. Their findings indicate that separating source and channel coding is optimal for maintaining communication concealment. Leveraging DNN-based techniques, they demonstrated that covert communication can be effectively realized, particularly for classification tasks. Notably, their study revealed that to ensure reliable JSCC while preserving covertness, the number of transmitted source symbols must scale with the square root of the number of channel uses.

1.4.5 Complementary strategies

This section explores methodologies for protecting SemCom that are distinct from the previously defined categories. The mitigation techniques discussed here either operate independently of these categories or address privacy leakage and eavesdropping issues in the fields similar to SemCom, with potential generalization to SemCom itself. For example, [69] investigates the wiretap channel in QAM modulation for SemCom, while [86] focuses on eliminating personal information in video transmissions. Additionally, [87], [88] propose content-based methods for protecting transmitted images. These studies provide valuable insights into privacy protection strategies that can be adapted or extended to enhance the security of SemCom systems.

In [69], the authors propose a DNN-based framework for semantic communication that extracts meaningful representations and compresses images while utilizing two distinct constellation maps for QAM modulation. This design aims to establish a nearly information-theoretically secure data transmission framework, enhancing both efficiency and security in SemCom systems. This method leverages superposition coding and employs a double-layered constellation map to integrate semantic features into satellite codewords. The MSE is employed as the loss function during the training phase, whereas in the power allocation process, the symbol error probability is evaluated to optimize transmission performance and security.

In the domain of privacy-preserving semantic video transmission, [86] introduces a novel decentralized Machine Learning (ML)-based video surveillance framework that leverages edge computing to protect personally identifiable information, particularly facial image data, by performing training at edge nodes. The authors address the security threats posed by Honest But Curious (HBC) attackers, who exploit their privileged access to extract sensitive information from transmitted video streams. To mitigate this risk, the proposed framework integrates privacy-preserving mechanisms that safeguard semantic information while ensuring both efficient and secure video transmission.

Beyond SemCom, alternative schemes exist that extract, transmit, and protect data privacy through different mechanisms. For instance, in [87], [88], the authors mitigate eavesdropping attacks by employing PLS and fountain-based packet delivery techniques for content-based image transmission, where the background and the Region of Interest (ROI) are encoded separately. While these approaches primarily focus on content-based processing without explicitly considering semantic meaning, we argue that they can be adapted and extended toward SemCom to enhance both security and efficiency in future communication systems.

1.5 Summary and Conclusion

This chapter has consolidated the main defense strategies proposed to counter eavesdroppers in SemCom. The existing methods are grouped into five categories: (a) techniques exploiting hardware or channel randomness (PLS); (b) encryption- and key management-based approaches, including quantum solutions; (c) data-driven methods that reshape semantic extraction and JSCC to balance the security–distortion trade-off; (d) semantic covertness approaches that conceal semantic content; and (e)

complementary strategies adapted from related domains or departing from conventional paradigms. Table 1.8 provides a holistic summary of these categories, outlining representative methods, their main strengths and limitations, and the lessons learned across different application scenarios.

PLS Methods PLS-based techniques extend conventional physical-layer security schemes, such as AN, beamforming, power allocation, and RIS, to the semantic domain. They effectively degrade Eve's reception performance while maintaining low computational complexity. However, most studies unrealistically assume full or partial knowledge of Eve's channel state information or location. In addition, current research is largely restricted to text-based communication, with limited exploration of image, video, or multimodal tasks. Future work should design PLS strategies robust to unknown or dynamic channels and capable of handling heterogeneous semantic data.

Encryption and Secret Key Management Encryption-based solutions provide strong theoretical protection by obfuscating transmitted signals prior to transmission. Works combining encryption, permutation, or quantum key distribution demonstrate promising adaptability to SemCom. Nonetheless, these methods often entail high computational and communication overhead, which limits their use in latency-sensitive or resource-constrained scenarios. Furthermore, quantization-based encryption may lead to semantic information loss, and lightweight schemes typically assume idealized secure key exchanges. Addressing the risk of key leakage remains an open challenge. Overall, encryption methods ensure robustness against eavesdropping regardless of Eve's channel or SNR, but must strike a balance between security strength, semantic fidelity, and practical efficiency.

Data-Driven Methods Data-driven techniques counter eavesdropping by optimizing loss functions that minimize Bob's reconstruction error while maximizing Eve's distortion. These methods are flexible, task-aware, and eliminate the need for explicit AN or encryption modules. By leveraging AI-based semantic extraction, particularly within JSCC frameworks, they offer natural adaptability to task semantics. However, several works rely on assumptions about Eve's channel or decoder, which may not be feasible in practice. Information bottleneck or sensitive-feature removal strategies relax these assumptions but require retraining for new tasks. Furthermore, adversarial training introduces additional complexity and may reduce reconstruction quality. Overall, data-driven approaches are promising due to their adaptability and semantic integration, though they still require enhancements in robustness and computational efficiency.

Semantic Covertness Semantic covertness represents an emerging line of defense, where protection is achieved by concealing or perturbing the transmitted meaning through techniques such as perturbation generation, steganography, or covert loss optimization. These methods inherently operate without relying on Channel State Information (CSI) or knowledge of Eve's model. However, steganographic designs based on deep architectures (e.g., CNNs) increase training and inference complexity,

while additional loss constraints may degrade reconstruction quality or task accuracy. Consequently, a trade-off exists between security and semantic fidelity at the receiver side. Despite these challenges, semantic covertness remains a promising paradigm—especially within JSCC—and future research should pursue lightweight, low-latency designs that preserve both security and meaning.

Complementary Strategies Complementary strategies encompass alternative or hybrid approaches, including double-layer constellation mapping, privacy-preserving video frameworks, and content-aware protection of sensitive regions. Each of these methods contributes unique mechanisms for securing semantic information. For example, the work in [86] introduces one of the few frameworks explicitly targeting video privacy. Extending such techniques to other modalities and integrating them with existing SemCom architectures remains an open opportunity. These methods often eliminate the need for key exchange, reducing the risk of key leakage, and offer new directions for semantic-level content protection. Further research should also address private attribute leakage in non-ROI regions and adapt these approaches to future multimodal SemCom systems. Collectively, complementary strategies broaden the security design space and highlight emerging opportunities for resilient, privacy-preserving semantic communication.

Eavesdropping in SemCom					
Defense methods	PLS methods • AN • Friendly jamming • Beamforming • RIS • Power allocation	Encryption and secret key management • Encryption (RSA,...) • Shuffling and permutation • Quantum security • Secret key management	Data-driven methods • Privacy-aware JSCC training • IB loss • Error measurement loss • Security-distortion trade-off	Semantic covertness • Covert loss • Steganography • Perturbation generation	Complementary strategies • Secure modulation • Privacy-preserving preprocessing at the edge • Content-based protection • PLS and fountain-based packet delivery
Pros / cons	🟢 Simple computations 🔴 Need for exact CSI and eavesdropper's information	🟢 Robust to channel changes 🟢 No Eve info required 🔴 High computational cost 🔴 Requires secure key exchange 🔴 Risk of key leakage 🔴 Possible quantization loss	🟢 No extra hardware 🟢 Tailored to JSCC in SemCom 🔴 High training cost 🔴 Needs retraining if the task changes	🟢 Robust to CSI changes 🟢 No Eve info required 🔴 Complex CNNs and training 🔴 Trade-offs may reduce task performance	🟢 No key exchange 🟢 Video privacy preserving 🔴 Non-ROI context may leak private attributes
Scenarios	Secure image / video / text tramsmission, UAVs, ISSC, quantum computing, single Eve, multiple Eves, multiple antennas, single-antenna, IoT, video surveillance				

TABLE 1.8: Summary of defense methods, their advantages, limitations, and application scenarios considered in the literature.

Chapter 2

Semantic-Aware Attention-Driven JSCC for Efficient Video Transmission over Wireless Channels

2.1 Introduction

The rapid growth of multimedia platforms and high-bandwidth wireless access has made video the dominant form of Internet traffic, accounting for nearly 80% of global traffic and expected to grow further [89]. This has driven research on efficient and robust wireless video transmission, traditionally based on source-channel separation. In these architectures, source coding (e.g., Advanced Video Coding (AVC)) compresses video into bitstreams, while channel coding protects them from wireless noise impairments. Despite their success, such designs are sensitive to channel mismatches: when actual conditions deviate from the assumed code rate, decoding errors can increase sharply, leading to complete reconstruction failures, known as the *cliff effect*. Furthermore, conventional schemes ignore semantic relevance and downstream task requirements, resulting in inefficient resource utilization [90], which is critical for emerging applications like wireless virtual and augmented reality requiring ultra-low latency with constrained devices [90].

Semantic communications have recently emerged as an end-to-end, task-oriented paradigm [14], transmitting only task-relevant information and potentially overcoming the limitations of conventional architectures. Deep learning techniques have proven effective for extracting and leveraging semantic features in multimedia transmission [9], [91], [92]. In particular, JSCC methods enhance robustness to channel variations, especially under low SNRs by providing graceful performance degradation while reducing latency and improving bandwidth efficiency [9]. Recent video-focused JSCC frameworks often adopt key and non-key frame structures [90], transmit latent features with or without explicit semantic priors [90], [93], [94], [95], and aim to exploit task-relevant information. While recent works have successfully integrated semantic information for video reconstruction, many existing frameworks rely on computationally intensive modules or prioritize the generative reconstruction of non-key frames. There remains a need for lightweight architectures that can robustly transmit key frames under variable channel conditions by fully exploiting compact semantic priors without incurring significant computational overhead.

The architecture integrates an adjacency matrix derived from segmentation maps, an SNR-aware modulation block (SNRMod), and residual blocks to jointly improve

semantic interpretability and reconstruction quality, while minimally increasing trainable parameters.

Motivated by these limitations, this work makes the following contributions:

- We propose a DL-enabled JSCC framework for key-frame transmission that integrates semantic attention (*SemAtt*) modules leveraging segmentation maps as priors to guide encoding and enhance feature discriminability. We focus on key frames since semantic errors at this stage can propagate and degrade the reconstruction of subsequent frames.
- We further introduce a novel Segmentation Adjacency Matrix (SGM)-based mechanism combined with SNR-aware modulation, improving semantic interpretability and reconstruction performance across varying channel conditions with negligible additional model complexity.
- Experimental results on benchmark datasets demonstrate high reconstruction fidelity under low SNR. Compared to conventional separation-based schemes prone to the digital cliff effect, our approach achieves a 5.2% reduction in LPIPS and a 6.3% improvement in Mean Intersection over Union (mIoU).

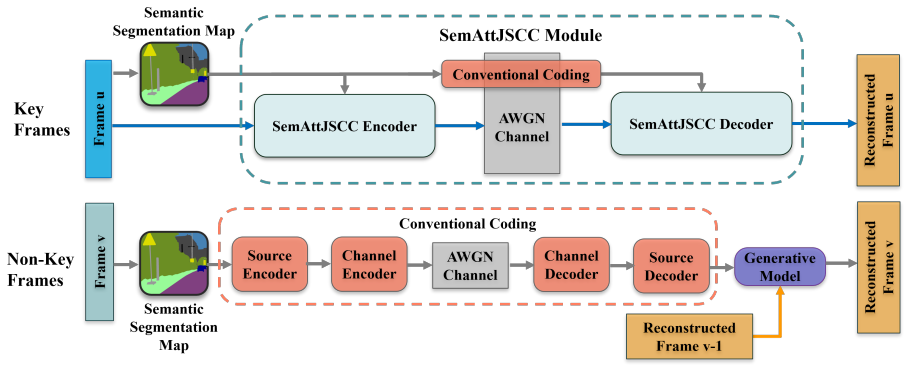

FIGURE 2.1: Semantic-aided JSCC framework overview.

The remainder of this chapter is organized as follows. Section 2.2 presents the proposed semantic-aided JSCC framework and details the SemAttJSCC module. Section 2.3 reports the experimental setup and performance evaluation against baseline JSCC and conventional coding schemes. Finally, Section 2.4 concludes the chapter

2.2 Semantic-Aided JSCC Framework for Video Transmission

This section introduces a semantic-aided JSCC framework for key-frame video transmission. Inspired by the object-attribute-relation paradigm in [94], the proposed architecture incorporates a *Semantic Attention* (*SemAtt*) mechanism that embeds semantic priors directly into the encoding and decoding stages.

2.2.1 Proposed Framework Architecture

Figure 2.1 illustrates the proposed framework. We consider wireless video transmission where sequences are divided into Group of Pictures (GoP), following [90]. The upper pipeline shows key-frame processing, with key frames defined as the first frame of each GoP. Semantic features are extracted from each key frame using pretrained HRNetV2 and Optical Character Recognition (OCR) models [96], [97], [98] to generate high-resolution segmentation maps. These maps, along with the RGB frame, are fed into the *SemAttJSCC* module, an end-to-end trainable system for joint source and channel coding. The module comprises a *SemAttJSCC Encoder* at the transmitter and a *SemAttJSCC Decoder* at the receiver. The encoder jointly processes the RGB image and its semantic map to generate complex-valued symbols transmitted over an Additive White Gaussian Noise (AWGN) channel. Given their compact representation, semantic maps are also transmitted via a parallel conventional source-channel coding pipeline, using 4-QAM modulation and a 1/3-rate Low-Density Parity Check (LDPC) code [94]. At the receiver, the decoded semantic information and key-frame symbols are jointly used by the decoder to reconstruct the RGB frame, with a symmetric architecture enabling end-to-end optimization.

The lower branch of the Figure 2.1, instead, corresponds to the non-key-frame processing pipeline within each GoP. To save bandwidth, only their semantic maps are transmitted via the separate coding scheme. At the receiver, the decoded map and the previously reconstructed RGB frame are fed into a vid2vid-based generative model [99] to synthesize the current frame. The *vid2vid* model is pretrained on the target dataset and deployed in inference mode at the receiver. As *vid2vid* operates exclusively on non-key frames, it does not impact the complexity analysis of key-frame transmission.

2.2.2 The SemAttJSCC Module

The proposed *SemAttJSCC* module, designed for key-frame transmission, extends the DynamicJSCC framework [100]. As shown in Figure 2.2, it comprises a jointly trained encoder-decoder pair: the *SemAttJSCC Encoder* and *Decoder*, each logically divided into source and channel blocks.

The source encoder E_S receives an RGB image $X \in \mathbb{R}^{3\times H\times W}$ and extracts high-level features, reducing spatial resolution by a factor G and expanding channel dimension to C, resulting in $X_S \in \mathbb{R}^{C\times H/G\times W/G}$. We adopt $G = 4$ and $C = 256$ as in [100]. The channel encoder E_C applies shape-preserving transformations, including a residual block (*ResBlock*) and an SNR-aware modulation block (*SNRMod*), to incorporate channel state information. A key novelty is the insertion of a *Semantic Attention* (*SemAtt*) block prior to *SNRMod* (highlighted in Figure 2.2), enabling semantic-aware feature modulation. Inspired by [101], [102], the *SemAtt* block exploits semantic segmentation priors via the SGM, guiding both attention and normalization stages.

The *SemAtt* block has three stages:

- **Attention layer:** processes latent features together with the SGM from a downsampled segmentation map.
- **Residual connection:** adds the attention output to the input features for stable gradient propagation [103].

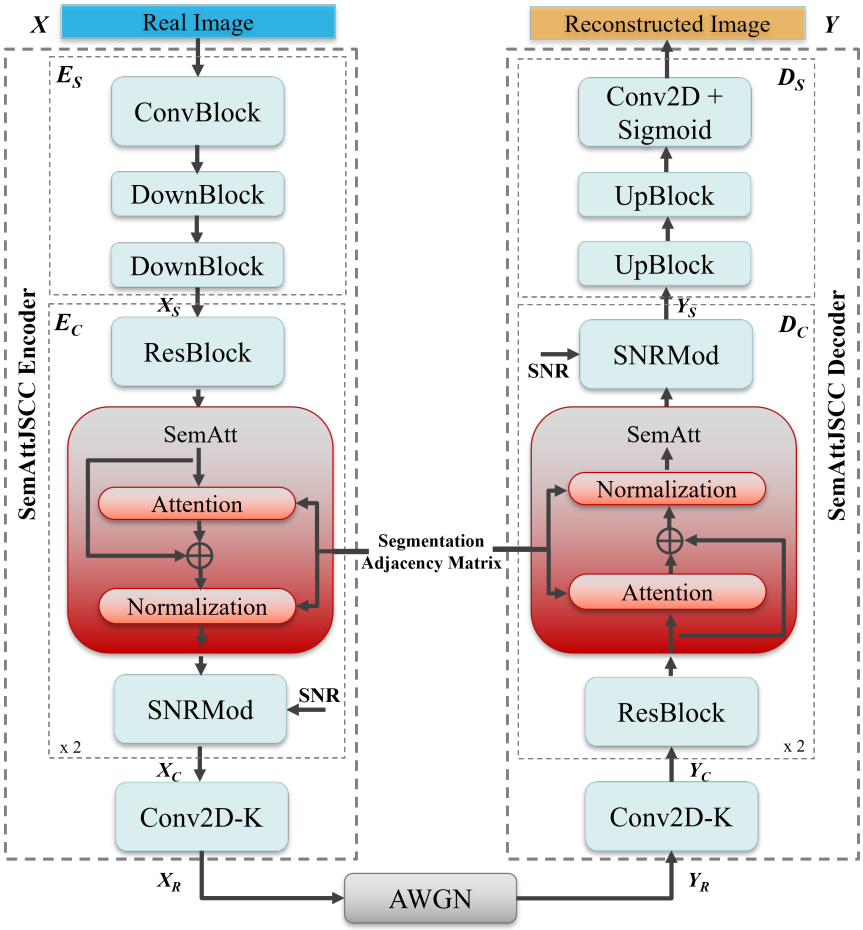

FIGURE 2.2: SemAttJSCC module architecture.

- **Semantic-aware normalization:** adjusts activations based on region-specific statistics.

The SGM $A \in \{0,1\}^{L\times L}$ is constructed from M semantic regions S_m, for $m = 1,\ldots,M$, where $L = H \times W$:

$$a_{ij} = \begin{cases} 1, & \text{if pixels } i,j \in S_m \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

To reduce memory and computation, a compressed representation $A_r \in \{0,1\}^{L\times M}$ is used, preserving semantic relationships with only L non-zero entries.

The feature tensor is reshaped into $F \in \mathbb{R}^{C\times L}$, and a deviation matrix D is computed as

$$d_{ij} = f_{ij} - \frac{\sum_{k=1}^{L} f_{ik}a_{kj}}{\sum_{k=1}^{L} a_{kj}}, \tag{2.2}$$

capturing channel-wise deviations from region-wise means. For each semantic region S_m, the attention map $\Psi^{S_m} \in \mathbb{R}^{C\times C}$ is defined by [102]:

$$\psi_{ij}^{S_m} = \frac{e^{\frac{1}{|S_m|}\sum_{k\in S_m} d_{ik}d_{jk}}}{\sum_{j=1}^{C} e^{\frac{1}{|S_m|}\sum_{k\in S_m} d_{ik}d_{jk}}}. \tag{2.3}$$

The attention output is equal to:

$$\varphi_{ij} = f_{ij} + \sum_{k=1}^{C} \psi_{ik}^{S_m(j)} f_{kj} \ , \tag{2.4}$$

followed by semantic-aware normalization:

$$z_{ij} = \frac{\varphi_{ij} - \mathrm{E}[\varphi_{ij}^{S_m(j)}]}{\sqrt{\mathrm{Var}[\varphi_{ij}^{S_m(j)}]}} w_{im} + b_{im} \ , \tag{2.5}$$

where $w, b \in \mathbb{R}^{C \times M}$ are learnable parameters. Assuming region-wise feature consistency, this reduces parameters by a factor M/L. The output is reshaped to $Z \in \mathbb{R}^{C \times H/G \times W/G}$ and passed to *SNRMod*. After *SNRMod*, a *Conv2D-K* layer projects features to K channels, followed by power normalization and mapping to complex symbols $X_R \in \mathbb{C}^k$, with $k = (K/2) \times H/G \times W/G$. The channel is modeled as AWGN:

$$Y_R = X_R + n \ , \quad n \sim \mathcal{CN}(0, \sigma^2 I_k), \tag{2.6}$$

with $\mathrm{SNR} = 10 \log_{10}(1/\sigma^2)\,\mathrm{dB}$.

The bandwidth compression ratio is $R = k/n$, with $n = 3HW$, controlled via K. At the receiver, Y_R is mapped back to a latent tensor and processed by the channel and source decoders D_C and D_S, mirroring E_C and E_S, to reconstruct Y. The module is trained end-to-end to minimize the reconstruction MSE:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \|X_i - Y_i\|^2 \ , \tag{2.7}$$

where N is the number of training samples.

2.3 Experimental Results

We evaluate the proposed JSCC framework for key-frame transmission on the *Cityscapes* dataset [104], a standard benchmark for semantic urban scene understanding. The dataset contains 2 975 training images and 500 validation images with resolution 1024×2048 and pixel-level annotations across 34 semantic classes.

The *SemAttJSCC* model is trained with the Adam optimizer [105] (learning rate 5×10^{-4}, batch size 2) on a single NVIDIA A100-SXM4-40GB GPU (CUDA 12.2). Early stopping with a patience of 14 epochs is applied, yielding convergence after roughly 35 epochs. During training, the SNR is uniformly sampled in $[0, 20]$ dB. Two bandwidth compression ratios are considered: $R = 1/12$ and $R = 1/6$. The DynamicJSCC baseline [100] is trained under identical conditions. Crucially, since this baseline already incorporates the *SNRMod* modulation mechanism, the performance comparison presented in this section effectively serves as an ablation study, isolating the specific contribution of the proposed *SemAtt* module.

The DynamicJSCC baseline [100] is trained under identical conditions. Experiments are limited to the AWGN channel and Cityscapes sequences to focus on the impact of compact semantic priors.

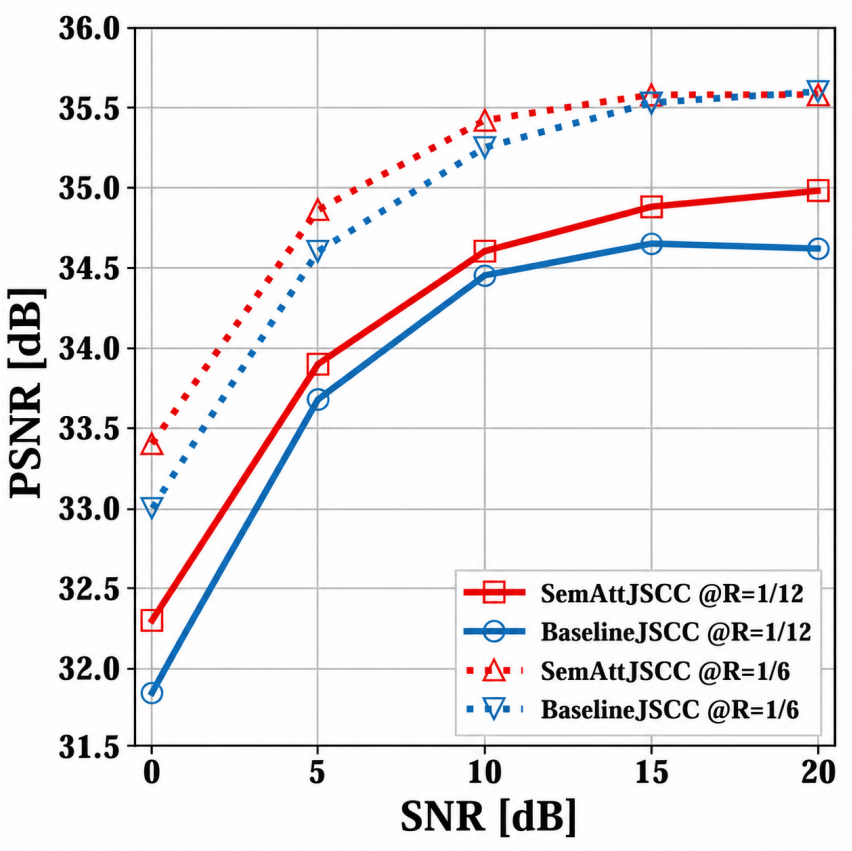

FIGURE 2.3: PSNR comparison between *SemAttJSCC* and the baseline for varying SNR and R.

2.3.1 Comparison with Baseline JSCC Scheme

Performance is evaluated using both fidelity-based and task-oriented metrics. PSNR and SSIM quantify pixel-level reconstruction quality, while LPIPS [106] measures perceptual similarity. Ground-truth segmentation maps are used to isolate transmission effects, and downstream semantic performance is evaluated via a pretrained HRNetV2+OCR model [96], [97], [107], computing mIoU on reconstructed images.

Figures 2.3 and 2.4 show that *SemAttJSCC* consistently outperforms the baseline across all SNR values and both compression ratios. Gains are more pronounced at stronger compression ($R = 1/12$) and low SNR, confirming improved robustness under adverse channel conditions. A maximum PSNR gain of 1.5% occurs at $SNR = 0$ dB and $R = 1/12$, while the largest SSIM improvement is 0.7% at $SNR = 0$ dB and $R = 1/6$.

Figure 2.5 shows that *SemAttJSCC* achieves consistently lower LPIPS scores than the baseline across all channel conditions, indicating superior perceptual quality. The maximum reduction is 13.7% at $SNR = 0$ dB and $R = 1/6$. Figure 2.6 demonstrates higher mIoU values, particularly under strong noise and compression, with the largest gain of 1.6% at $R = 1/12$ and $SNR = 5$ dB. These results confirm that semantic-aware encoding effectively preserves class-discriminative information for downstream tasks.

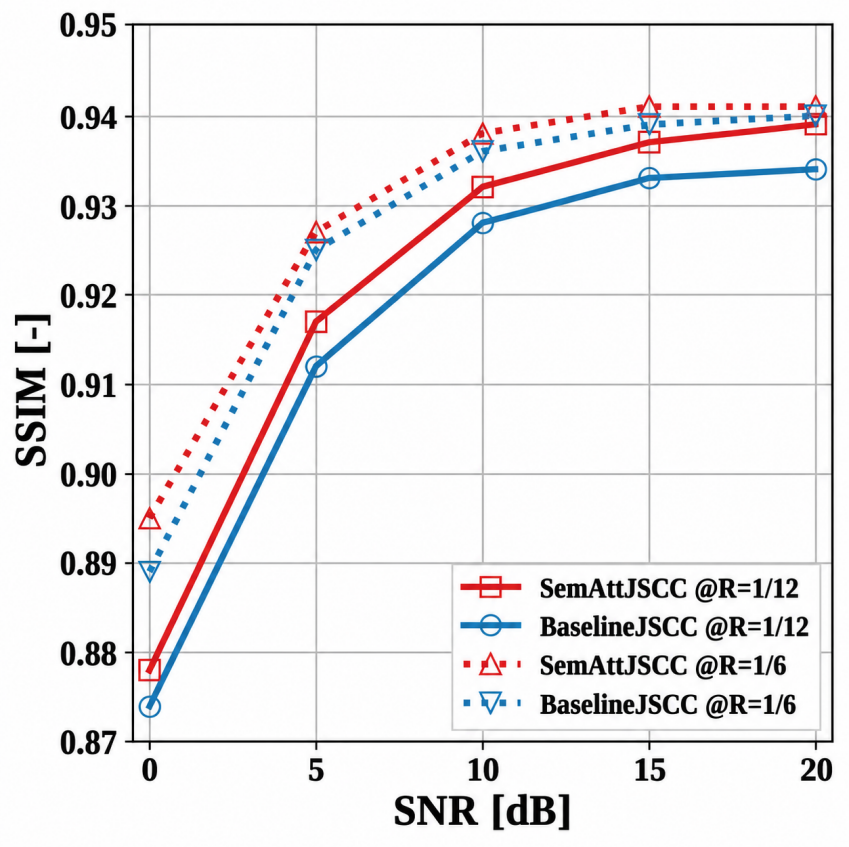

FIGURE 2.4: SSIM comparison between *SemAttJSCC* and the baseline for varying SNR and R.

2.3.2 Comparison with Conventional Coding Schemes

We further evaluate the proposed *SemAttJSCC* framework against conventional separate source and channel coding pipelines, following [94]. In these reference schemes, images are first compressed using JPEG2000 or BPG [108], then protected via LDPC coding [109] and digitally modulated. We consider LDPC code rates of 1/3, 1/2, and 2/3, with BPSK, 4-QAM, 16-QAM, and 64-QAM modulation. Perfect CSI is assumed at the receiver for Log-Likelihood Ratio (LLR) computation [110]. For fair comparison, the compressed file sizes are matched to the number of transmitted symbols used by *SemAttJSCC*, including the overhead associated with transmitting the compressed SGM. It is worth noting that this overhead is minimal, amounting on average to 0.4% of the compression bandwidth ratio R.

Figures 2.7 and 2.8 highlight that JSCC-based methods inherently avoid the digital cliff effect typical of conventional digital pipelines. In terms of perceptual quality (measured via LPIPS), *SemAttJSCC* consistently outperforms JPEG2000-based schemes across all SNR values. Compared to BPG-based pipelines, the proposed method achieves superior reconstruction fidelity in low-SNR regimes, where digital schemes often fail due to unrecoverable bit errors. Specifically, at $SNR = 0$ dB, *SemAttJSCC* reduces LPIPS by 5.2% and increases mIoU by 6.3% relative to the BPG+1/3LDPC+BPSK configuration. While BPG demonstrates competitive performance at high SNR due to efficient source coding, this advantage primarily stems from optimized compression rather than transmission robustness, which is the focus

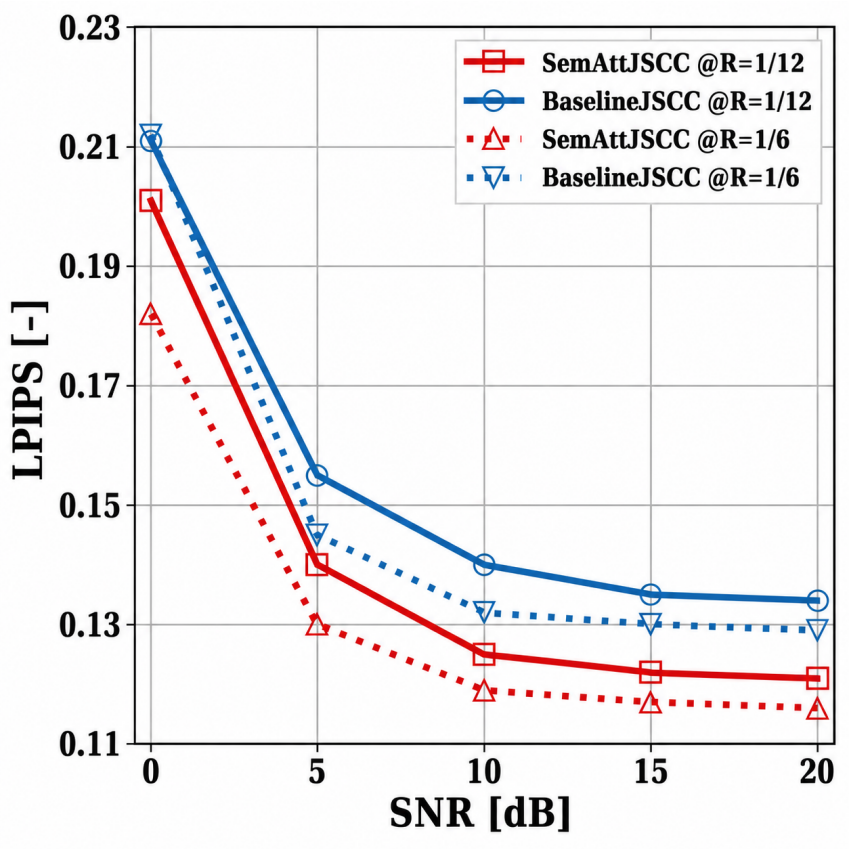

FIGURE 2.5: LPIPS comparison between *SemAttJSCC* and the baseline for varying SNR and R.

of our study. Additionally, the high-resolution images (1024×2048) pose additional challenges for conventional and learned methods [94], [110]. Future work could explore increasing the latent channel dimensionality K to better balance spatial compression and feature richness under a fixed bandwidth constraint. Similar trends are observed for downstream semantic tasks. Figure 2.8 shows that *SemAttJSCC* significantly outperforms both JPEG2000- and BPG-based pipelines in terms of mIoU, particularly at low SNR, confirming the effectiveness of semantic-aware encoding in preserving task-relevant information. A qualitative example is provided in Figure 2.9, showing reconstructed key frames and corresponding segmentation outputs

TABLE 2.1: Computational complexity and number of trainable parameters for the proposed *SemAttJSCC*, the baseline JSCC, and the convolution-enhanced JSCC model [94], evaluated at bandwidth compression ratio $R = 1/12$ for key-frames transmission.

Architecture	**GFLOPs**	**Parameters (M)**	**Max. LPIPS Reduction vs. Baseline**
Baseline JSCC	2079	6.57	–
SemAttJSCC	2285	6.64	13.7%
Baseline JSCC + Convolutions [94]	3326	11.33	11.0%

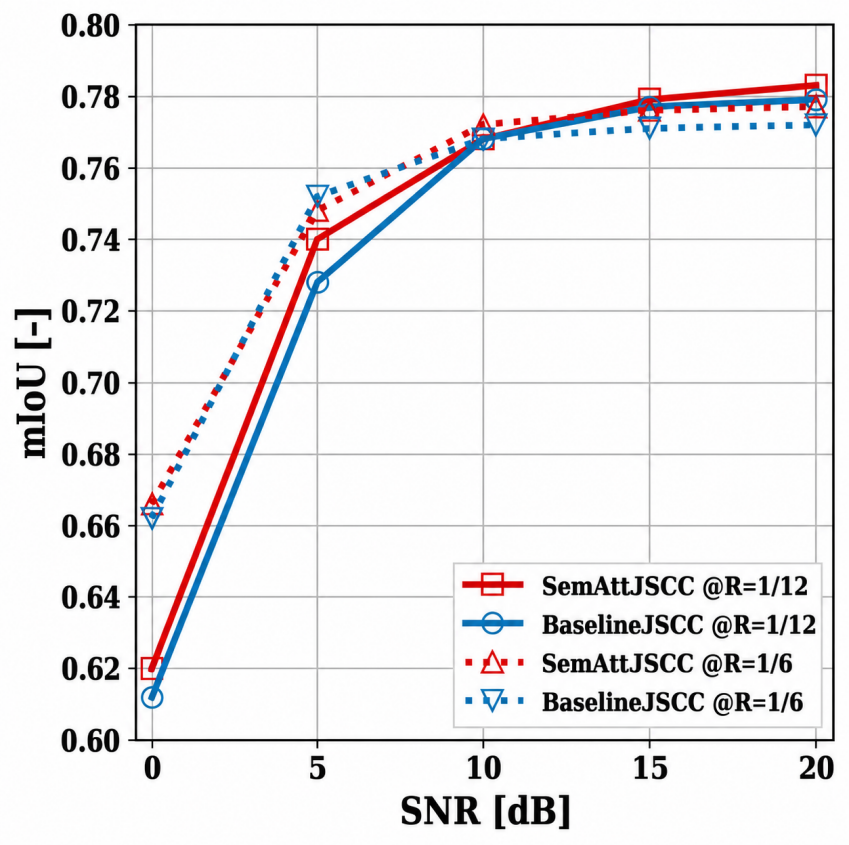

FIGURE 2.6: mIoU comparison between *SemAttJSCC* and the baseline for varying SNR and R.

at $SNR = 0$ dB. The proposed method better preserves fine structures and semantic boundaries compared to conventional schemes, even under severe channel noise. In addition to improved performance, *SemAttJSCC* introduces only modest computational overhead. As summarized in Table 2.1, GFLOPs and trainable parameters increase by only 9.9% and 1.1%, respectively, relative to the baseline JSCC. By contrast, the convolution-enhanced model in [94] incurs substantially higher overhead (60.0% GFLOPs and 72.5% parameters) while achieving a smaller maximum LPIPS reduction (11.0% vs. 13.7%). These results, albeit obtained on datasets of different resolutions, underscore the efficiency of integrating compact semantic priors directly within the JSCC framework.

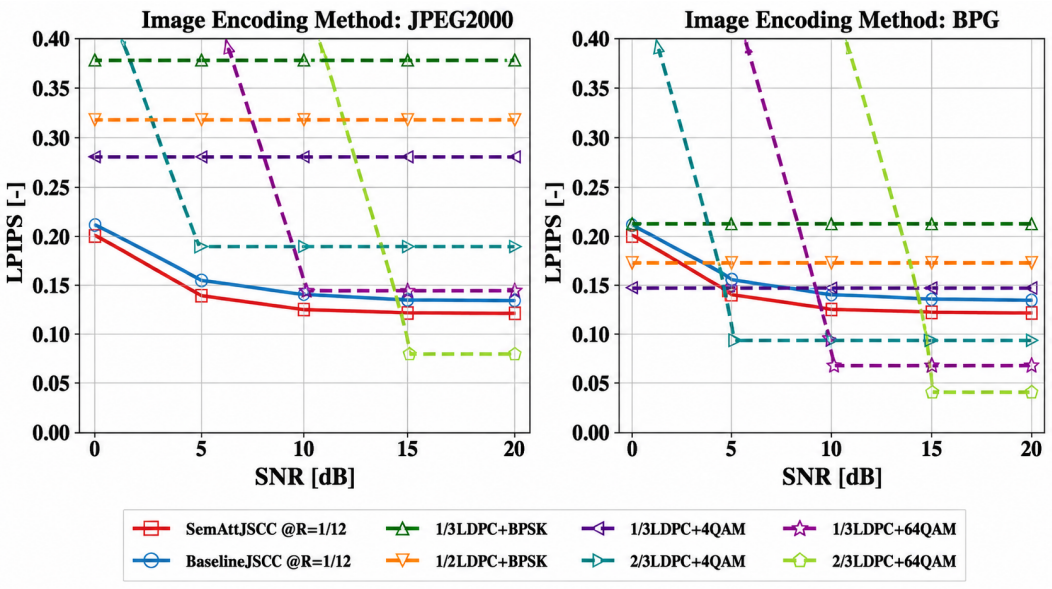

FIGURE 2.7: LPIPS comparison with conventional coding schemes under varying SNR values.

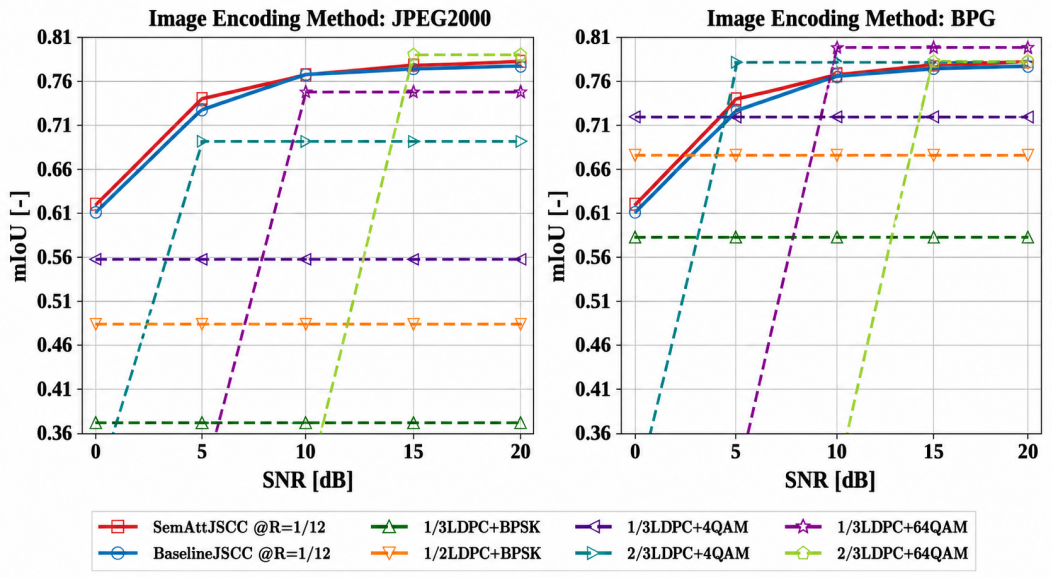

FIGURE 2.8: mIoU comparison with conventional coding schemes under varying SNR values.

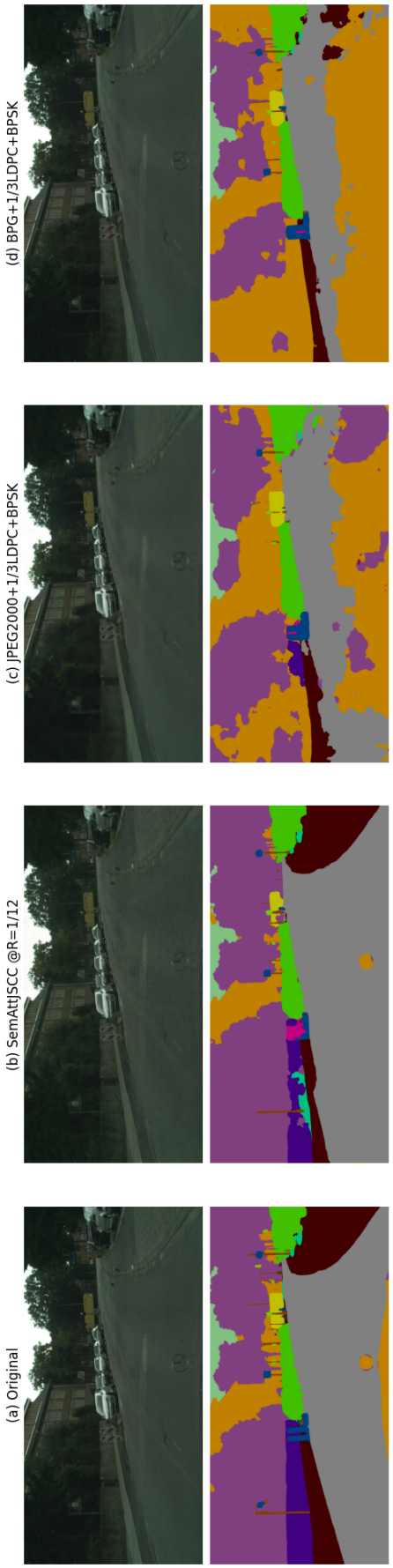

FIGURE 2.9: Visual comparison of reconstructed key frames (top row) and corresponding segmentation outputs (bottom row) at $SNR = 0$ dB.

2.4 Conclusion

We introduced *SemAttJSCC*, a semantic-aware DJSCC framework for efficient key-frame transmission over bandwidth-limited, noisy wireless channels. By integrating compact semantic priors (SGMs) via lightweight attention mechanisms, the system guides feature allocation and channel adaptation. Experiments on high-resolution urban scenes show consistent gains over baseline JSCC in reconstruction (PSNR, SSIM), perceptual quality (LPIPS), and downstream semantic accuracy (mIoU), particularly under low SNR and strong compression. Compared to conventional digital pipelines (JPEG2000/BPG + LDPC), *SemAttJSCC* avoids the cliff effect and maintains semantic integrity. Future work will explore more realistic channel models, adaptive GoP structures, key-frame selection under strict latency and bandwidth constraints, and generative schemes for reconstructing non-key frames from semantically transmitted key frames.

Chapter 3

Eavesdropping and Model Inversion Attacks in SemCom over NTNs

3.1 Introduction

Shannon's classical information theory has long underpinned communication systems by ensuring bit-level fidelity, yet it remains agnostic to the semantic content of transmitted data. SemCom introduces a paradigm shift beyond Shannon's model by enabling the transmission of meaning-relevant information, thereby enhancing communication efficiency. By leveraging AI, SemCom prioritizes task-relevant features, achieving significant reductions in bandwidth usage and latency, while improving throughput and system performance. Despite these advantages, SemCom introduces novel security vulnerabilities. The inherent openness of wireless channels, high correlation between source and transmitted representations, and structural weaknesses in AI models render SemCom particularly susceptible to eavesdropping and privacy breaches. These risks necessitate a reassessment of traditional security paradigms. This chapter provides a comparative analysis of existing strategies under model inversion-based eavesdropping attacks, offering a unique perspective absent in existing literature. Then, a new method "KAP-DJSCC" for providing a secure SemCom is presented. The section concludes by identifying critical research challenges and outlining prospective directions for secure and trustworthy SemCom systems.

The remainder of this chapter is structured as follows: Section 3.2 focuses on MIEA, a specialized form of eavesdropping, and analyzes representative strategies designed to mitigate such threats. Section 3.3 proposes a new method (KAP-DJSCC) for overcoming eavesdroppers in JSCC. Open challenges and promising future research directions are discussed in Section 3.4. Finally, the conclusion is presented in Section 3.5.

3.2 MIEA

This section investigates the security threat posed by the MIEA, as introduced in [37], and its implications for JSCC. The attack is first described, followed by a comparative analysis of Eve's capability to reconstruct the received stream transmitted via JSCC. This is investigated in four scenarios: no protection, data-driven protection, lightweight cryptographic protection, and a hybrid approach.

3.2.1 MIEA Overview

Since SemCom fundamentally relies on AI techniques, it introduces novel security vulnerabilities absent in conventional communication systems. Notably, adversaries may exploit internal feature representations within semantic models to reconstruct the original data by inverting the semantic extraction process. This class of threats is encapsulated in the MIEA, which poses a significant privacy risk in SemCom deployments. CNN-based architectures are particularly susceptible to MIEA due to the high correlation between input data and the learned feature embeddings. In this work, we investigate MIEA within one of the most widely adopted CNN-based frameworks, the JSCC, which has demonstrated notable advancements in SemCom, especially for image compression and wireless transmission.

Unlike traditional communication schemes that perform source and channel coding separately, JSCC integrates both operations into a unified end-to-end learning framework. This joint optimization confers several advantages, including improved reconstruction fidelity, increased spectral efficiency, and inherent robustness against the *cliff effect*—a sharp degradation in performance typically observed in classical systems under marginal channel deterioration [9], [42]. Nevertheless, by replacing conventional communication blocks with DL components, JSCC becomes inherently vulnerable to MIEA, primarily due to the broadcast nature of wireless channels and the architectural sensitivity of CNN-based systems [111]. Consequently, a warden may intercept the wireless signal and accumulate sufficient data to train a surrogate decoder capable of inferring the semantic content or reconstructing the transmitted data.

In a typical MIEA scenario, the attacker seeks to inject inputs into the sender's network and simultaneously intercept the corresponding wireless transmissions via a dedicated eavesdropping channel. By collecting a sufficiently large set of intercepted data, the adversary trains a neural network to approximate the functionality of the legitimate semantic decoder, thereby recovering the original information.

Let $\mathbb{B} = \{b_1, b_2, \ldots, b_i, \ldots, b_S\}$ denote the training dataset used by the JSCC system, where S is the number of samples. The autoencoder comprises Alice's encoder E_θ and Bob's decoder D_β, with θ and β representing their respective parameter sets. For image transmission, the encoder E_θ transforms a raw input image x into a semantic latent representation and subsequently into transmittable symbols. The receiver (Bob) receives the noisy signal $y_B = E_\theta(x) + N_B$, where $N_B \sim \mathcal{CN}(0, \sigma_B^2)$ denotes the AWGN affecting Bob's channel. Similarly, the attacker (Eve) captures the signal via her own channel as $y_E = E_\theta(x) + N_E$, where $N_E \sim \mathcal{CN}(0, \sigma_E^2)$ models her channel noise. To perform the attack, Eve constructs her training dataset $\mathbb{E} = \{e_1, e_2, \ldots, e_j, \ldots, e_M\}$, where each input sample e_j is encoded using the known encoder E_θ to generate features. She simultaneously intercepts the corresponding noisy features y_E through her own channel. After collecting sufficient data, Eve trains her inverse decoder D_Ψ, parameterized by Ψ, to map the intercepted features back to the semantic content. The success of Eve's reconstruction depends on several factors, including the SNR of her channel, the structure and training quality of her decoder, and her prior knowledge about the encoder. This includes knowledge of the encoder's architecture, training weights, and data distribution. Accordingly, MIEA can be studied under varying conditions, such as whether Eve has access to the encoder's trained neural network model and parameters (white-box) or does not know anything about

the model, and only the output is provided (black-box). Without loss of generality, we refer to the encoder's model parameters as *weights*. In this chapter, we consider a partially white-box MIEA setting [112], [113], where Eve may have access to the encoder's weights and/or the training data, and we systematically evaluate how this knowledge impacts her reconstruction performance.

3.2.2 Assumptions and Mitigation Methods Against MIEA

As discussed in the previous section, the effectiveness of a MIEA largely depends on the adversary's prior knowledge and system access. Different assumptions regarding the attacker's capabilities lead to varying levels of reconstruction quality. We consider four distinct attack scenarios that reflect increasing levels of knowledge and access available to the eavesdropper:

1. **Unknown weights and dataset:** Eve has knowledge of the encoder's architecture but lacks access to the trained encoder weights and the original training dataset $\mathbb{B}$. In this scenario, Eve randomly initializes her decoder D_Ψ and trains it using an independently generated dataset $\mathbb{E}$, while receiving the intercepted features y_E from her eavesdropping channel.

2. **Known weights, unknown dataset:** Eve is aware of both the architecture and weights of the legitimate encoder E_θ, but still trains her decoder D_Ψ using a dataset $\mathbb{E} \neq \mathbb{B}$, introducing a distributional mismatch in the reconstruction process.

3. **Unknown weights, same data distribution:** Eve does not have access to the encoder weights, which she initializes randomly, but she possesses a dataset $\mathbb{E}$ drawn from the same distribution as $\mathbb{B}$, thus aligning the semantic context of training despite model disparities.

4. **Full access:** In this idealized and least likely case, Eve is assumed to have complete knowledge of the encoder architecture, its weights, and the training dataset. She performs training of her decoder D_Ψ using the exact distribution $\mathbb{B}$, thereby maximizing her capacity for accurate semantic reconstruction.

To defend JSCC systems against MIEA under the aforementioned scenarios, we propose and evaluate three mitigation strategies, each designed to impair the adversary's reconstruction capabilities while preserving the performance of the legitimate receiver:

- **Data-driven adversarial training:** Inspired by the method in [46], this approach modifies the training objective to minimize the MSE between the legitimate receiver's reconstruction and the original input, while simultaneously maximizing the MSE between the adversary's output and a predefined target, such as a black image. In contrast to [46], where Eve's SNR is assumed to be lower than Bob's, we adopt a more conservative setting by assuming a higher SNR for Eve, thereby evaluating the defense under a worst-case eavesdropping condition.

- **Lightweight cryptographic permutation and substitution:** Based on [37], this technique applies a cryptographic transformation to the transmitted semantic features by performing random permutation and substitution operations. While the original method requires transmitting two tensors (semantic features and noise) per image, our refined approach operates directly on the latent dimensions of the transmitted feature vector. The cryptographic key is assumed to be securely shared between Alice and Bob. This reduces communication overhead while maintaining effective protection against semantic inversion.

- **Hybrid defense approach:** This method combines the advantages of data-driven adversarial training and lightweight cryptographic schemes. Specifically, the latent semantic representation used during adversarial training is further obfuscated via permutation and substitution, both during training and inference. This dual-layer strategy significantly impairs the attacker's ability to learn an accurate inverse mapping, even under favorable assumptions.

All three methods aim to mitigate MIEA by either distorting the semantic representation extracted by the DNN or encrypting the feature stream before transmission. By weakening the correlation between the original input and the transmitted features, these countermeasures hinder the ability of unauthorized networks to reconstruct or infer sensitive semantic content.

3.2.3 Environment Setup

This section outlines the experimental framework adopted to evaluate the effectiveness of the proposed mitigation strategies against MIEA. Initially, the CIFAR-10 dataset [114] is employed to analyze information leakage under the MIEA scenario. The CIFAR-10 dataset is partitioned into 40,000 samples for training Bob's autoencoder, 10,000 for testing, and 10,000 for training the Eve's decoder D_{Ψ}. The learning rate is set to 0.0001, and the final convolutional layer of the encoder is configured with a depth of 8. The low-resolution CIFAR-10 dataset is selected because it is widely adopted in SemCom security analysis due to its balanced complexity and standardized benchmarking capability [8], [57]. Furthermore, to demonstrate the generalization capability of the proposed approach, we extend our evaluations to the SCUT-FBP5500 dataset, which comprises more complex, higher-resolution, and specific images containing human faces. The SCUT-FBP5500 dataset includes a total of 5,500 samples, which are divided into 3,465 images for training Bob, 1,485 for training Eve, and 550 for testing. This extension enables assessment of model robustness across datasets with varying semantic and visual complexities.

For the data-driven privacy-aware method inspired by [46], the SNR of Eve's channel is treated as independent from that of the legitimate receiver. Following the original methodology, a simulated adversarial decoder, architecturally identical to Bob's decoder, is incorporated into the loss computation during training. This simulated Eve does not represent a real adversary but instead enables the encoder to optimize the privacy–utility trade-off under realistic conditions. In contrast, the real Eve employs an independent decoder D_{Ψ}, trained separately through a MIEA.

The adversarial training incorporates the *SecureMSE* loss function proposed in [46], which consists of two MSE terms: one measuring the reconstruction fidelity

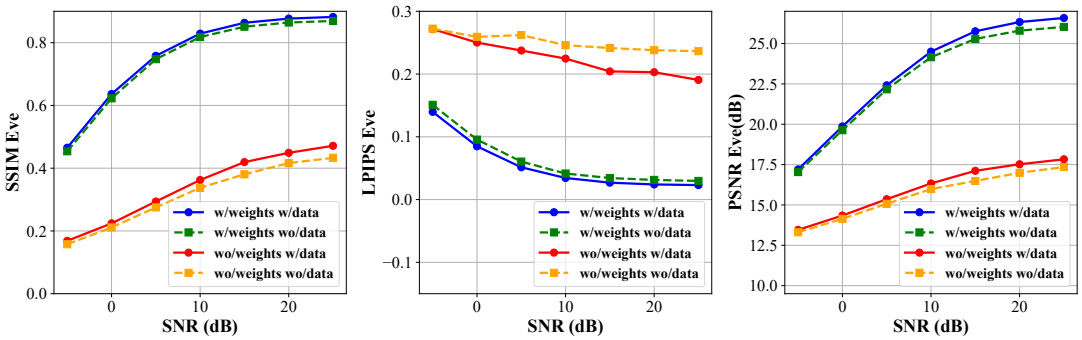

FIGURE 3.1: Eve's reconstruction quality under various access conditions without protection using CIFAR-10 dataset.

of the legitimate receiver, and another penalizing semantic similarity between Eve's output and a black reference image. This dual-objective optimization encourages the encoder to retain fidelity for Bob while suppressing meaningful information for Eve. To emulate a realistic and challenging attack environment, Bob is trained under a channel SNR of 15 dB, whereas Eve operates at a higher SNR of 20 dB.

Regarding the lightweight cryptographic strategy based on [37], the original JSCC architecture from [9] is adopted. This includes the encoder and decoder modules without modification. In the case of the hybrid defense method, the same training and architectural configurations as in the data-driven approach are applied, with the added integration of permutation and substitution operations. Following the training phase, MIEA is conducted to evaluate each defense strategy. The adversary's decoder D_Ψ is constructed using one upsampling layer followed by two convolutional layers, as described in [37]. For each SNR level in the set {-5, 0, 5, 10, 15, 20, 25} dB, Eve trains her decoder using corresponding semantic representations and samples drawn from her training dataset.

To quantify the quality of image reconstruction at both the legitimate receiver and the adversary, we employ three widely accepted metrics: PSNR, SSIM, and LPIPS, as summarized in Table 1.3. The pixel-level metrics, PSNR and SSIM, are formally defined as follows:

$$\text{PSNR}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \left(\frac{255^2}{\text{MSE}(\mathbf{x}, \hat{\mathbf{x}})} \right) \text{ [dB]}, \tag{3.1}$$

$$\text{SSIM}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + C_1)(2\sigma_{\mathbf{x}\hat{\mathbf{x}}} + C_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}^2 + C_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + C_2)}, \tag{3.2}$$

where $\mathbf{x}$ and $\hat{\mathbf{x}}$ denote the original and reconstructed images, $\mu_{\mathbf{x}}$ and $\mu_{\hat{\mathbf{x}}}$ represent their mean intensities, and $\sigma_{\mathbf{x}}^2$, $\sigma_{\hat{\mathbf{x}}}^2$ are their variances. The covariance $\sigma_{\mathbf{x}\hat{\mathbf{x}}}$ captures structural similarity, and C_1, C_2 are stabilization constants [115]. In addition, LPIPS [116] is used to assess perceptual quality. Unlike pixel-wise metrics, LPIPS evaluates high-level feature similarity using deep neural networks. In this study, LPIPS is computed using the pre-trained AlexNet [117], providing a measure that correlates strongly with human visual perception.

3.2.4 Simulation Results

This subsection investigates Eve's ability to reconstruct transmitted content under the scenarios introduced in Subsections 3.2.2 and 3.2.3. Additionally, a comprehensive comparison is conducted between existing defense mechanisms against MIEA and the approach proposed in this work. The impact of these defense strategies on Bob's performance is also examined.

Evaluation on CIFAR-10 dataset

Figure 3.1 presents the quality of images reconstructed by Eve in the absence of protection, across the four scenarios defined in Subsection 3.2.2. Metrics such as SSIM, PSNR, and LPIPS are evaluated over a range of Eve's channel SNRs. The results underscore the critical influence of model weight and data distribution access on Eve's reconstruction success. Specifically, the blue and green curves (with weight access) significantly outperform the orange and red ones (without weight access), demonstrating the heightened risk posed by weight leakage. Quantitatively, access to both model weights and data distribution nearly doubles the SSIM, increases PSNR by approximately 56.37%, and reduces LPIPS by around 90%. Table 3.1 summarizes the peak reconstruction performance of Eve across all SNR levels relative to the baseline case (no access).

TABLE 3.1: Improvement in Eve's reconstruction performance due to access to weights and data, CIFAR-10 dataset.

Access Scenario	**SSIM (%)**	**LPIPS (%)**	**PSNR (%)**
wo/weights w/data	10.22	3.81	19.37
w/weights wo/data	194.99	53.40	87.43
w/weights w/data	201.46	56.26	90.25

Notably, in the low-SNR regime, Eve's reconstruction performance remains poor in all scenarios due to high noise levels and mismatched operational and training SNRs. However, as Eve's channel SNR increases, particularly up to 15 dB, the reconstruction quality improves sharply, owing to reduced channel noise and improved alignment with Bob's conditions. Beyond 15 dB, the performance trend saturates, consistent with standard JSCC behavior [9].

Figure 3.2 shows a randomly selected image reconstructed by Eve under the four defined access conditions at 15 dB channel SNR, without any defense applied. The visual degradation in the absence of weight access confirms that random initialization significantly hinders reconstruction. This highlights the inherent privacy offered by SemCom, as accurate recovery of either model weights or data distribution is nontrivial under MIEA, even at favorable SNR. When Eve possesses full system knowledge, reconstruction quality is markedly higher, with weight access contributing more substantially than data alignment.

Applying defense mechanisms significantly impairs Eve's reconstruction capability. Figure 3.3 illustrates Eve's reconstructed outputs under various protection methods, assuming access to the model weights but not the data distribution. At both -5 dB and 20 dB SNR, the data-driven approach inspired by [46] leads to predominantly dark outputs, though some structure remains. In contrast, the lightweight

cryptographic method based on [37] effectively scrambles image features, distorting textures and preserving only low-frequency texture (e.g., sky or ground). The proposed hybrid strategy, which combines both approaches, offers superior protection by simultaneously disrupting pixel-level and semantic information.

FIGURE 3.2: Eve's reconstructions under different access conditions for a random image from CIFAR-10 dataset ($SNR_e = 15$ dB, no defense).

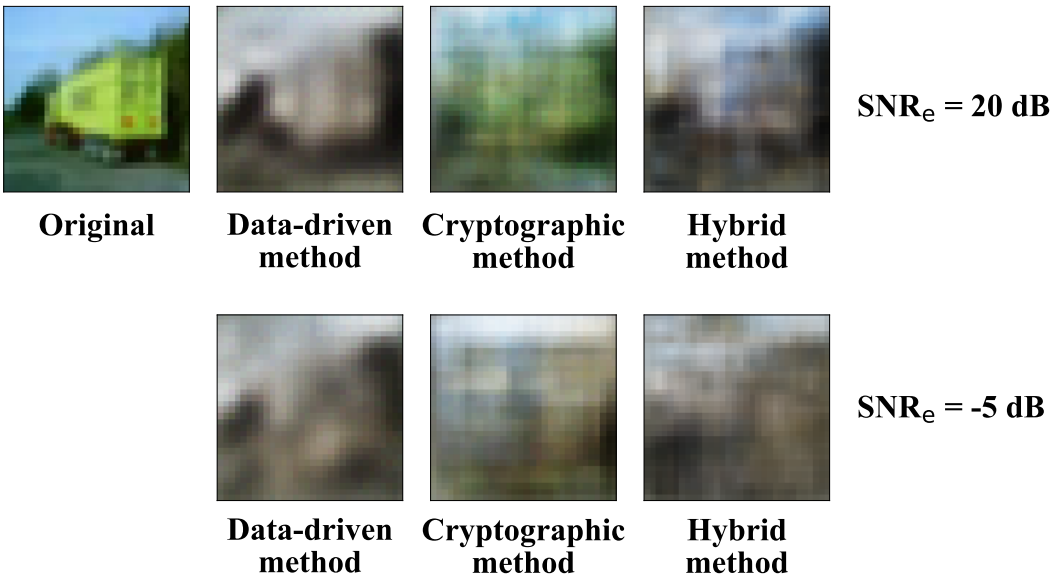

FIGURE 3.3: Eve's reconstructions under protection mechanisms with weight access but no data for a random image from CIFAR-10 dataset.

Table 3.2 summarizes the performance and computational complexity of the defense methods considered applied to the CIFAR-10 dataset. Complexity is reported in Giga Floating-point Operations Per Seconds (GFLOPs), excluding the cost of secure key transmission. The cryptographic approach is computationally lightweight, relying only on permutation operations. The data-driven and hybrid methods incur higher computational costs due to additional network components. Furthermore, Eve's reconstruction quality is evaluated under two extremes: full access to weights and data, and no access to either. The results confirm that access to system knowledge significantly boosts Eve's capabilities. However, even in the unprotected case, Eve performs worse without access than with protection in place. Among the strategies, the hybrid method offers the best protection, whereas the cryptographic method ensures minimal system overhead with strong privacy guarantees.

Method		**Eve ($SNR_e = 15$ dB)**					**Bob ($SNR_b = 15$ dB)**		
Method	**Complexity (GFLOP)**	**Weight Access**	**Data Access**	**SSIM ↓**	**LPIPS ↑**	**PSNR ↓**	**SSIM ↑**	**LPIPS ↓**	**PSNR ↑**
None	0.0278	✓ ✗	✓ ✗	0.86 0.38	0.02 0.24	25.76 16.49	0.91	0.008	27.72
Data-driven [46]	0.0365	✓ ✗	✓ ✗	0.59 0.35	0.30 0.29	19.62 16.05	0.81	0.07	24.20
Cryptographic [37]	0.0278	✓ ✗	✓ ✗	0.50 0.26	0.18 0.30	18.99 15.32	0.91	0.008	27.71
Hybrid	0.0365	✓ ✗	✓ ✗	0.46 0.25	0.20 0.28	18.17 15.25	0.61	0.12	19.94

TABLE 3.2: Eve's and Bob's reconstruction quality under different defense strategies and access scenarios, along with their corresponding computational costs evaluated on the CIFAR-10 dataset.

Evaluation on the SCUT-FBP5500 Dataset

Figure 3.4 illustrates Eve's reconstruction performance under the proposed Hybrid protection strategy, assuming Eve has full access to the encoder. Both Bob and Eve are trained and tested on the SCUT-FBP5500 dataset, with the training and testing partitions defined earlier in this section. In this setup, Bob's training SNR is fixed at 15 dB. The figure demonstrates how Eve's reconstruction capability varies as her training SNR differs from the testing SNR, thereby simulating realistic channel variations that an eavesdropper may encounter in practice. The evaluation metrics, including PSNR, SSIM, and LPIPS, reveal that under MIEA, both Bob's optimized autoencoder parameters and Eve's inaccurate estimation of the channel conditions affect the extent of recoverable information. Consistent with the observations in [9], Eve exhibits a graceful degradation in performance when $SNR_{test} < SNR_{train}$, avoiding the sharp "cliff effect" observed in conventional digital systems. Nevertheless, even at high SNRs, where channel noise is negligible, Eve fails to accurately reconstruct the transmitted content. This failure occurs primarily because the Hybrid protection method effectively conceals semantic representations through adversarial regularization and latent feature permutation. Moreover, the mismatch between Eve's training and testing SNRs prevents her decoder from adapting to the actual channel distribution. This analysis, together with Table 3.2, demonstrates that the proposed protection mechanism significantly reduces Eve's reconstruction capability and generalizes well across datasets of varying complexity.

Figure 3.5 compares Eve's reconstruction performance on the SCUT-FBP5500 dataset under various AWGN channel conditions, considering the defense strategies described in Subsection 3.2.2. The results indicate that the data-driven approach leaks the most information across all evaluation metrics, as it neither explicitly obfuscates the transmitted data nor employs adversarial perturbations to mislead Eve. In contrast, the proposed Hybrid method achieves up to a 58% reduction in Eve's recovery quality relative to the data-driven approach (based on the SSIM metric), while simultaneously increasing Eve's LPIPS by 44.5% and decreasing her PSNR by 30%. These improvements confirm the Hybrid method's superior ability to degrade Eve's perceptual and structural reconstruction accuracy.

Additionally, Figure 3.6 presents a visual comparison of Eve's reconstructed images for a randomly selected test sample. The cryptographic method effectively obscures the transmitted information; however, at higher SNR values (e.g., 25 dB), certain structural cues, such as borders and background textures, may still be partially preserved. The data-driven method tends to darken Eve's reconstructions while

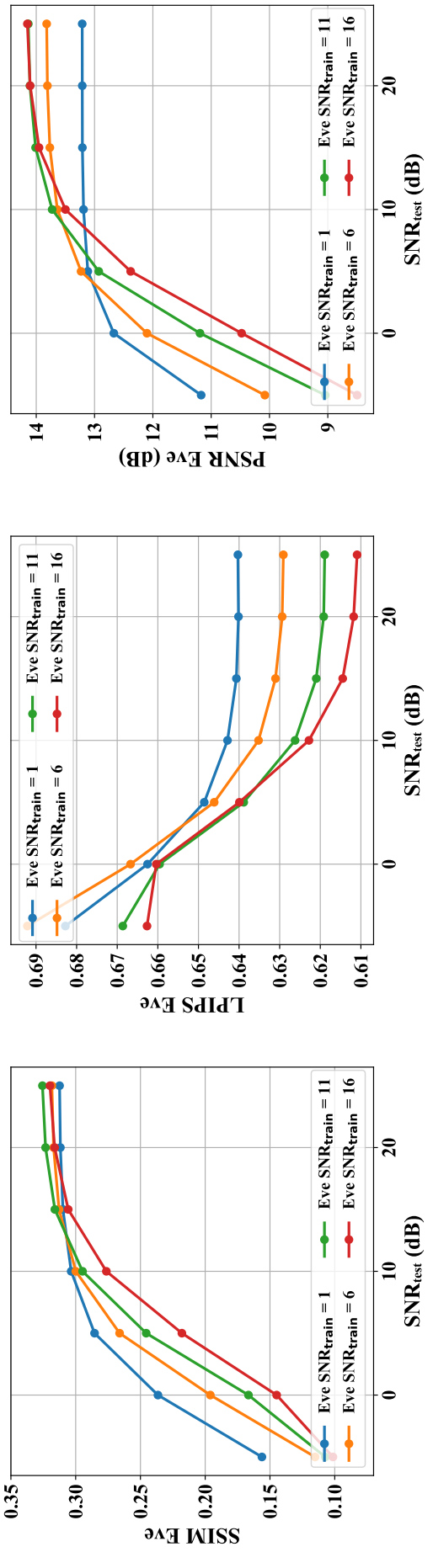

FIGURE 3.4: Eve's reconstruction quality under different training and testing SNRs when the Hybrid protection method is applied, SCUT-FBP5500 dataset.

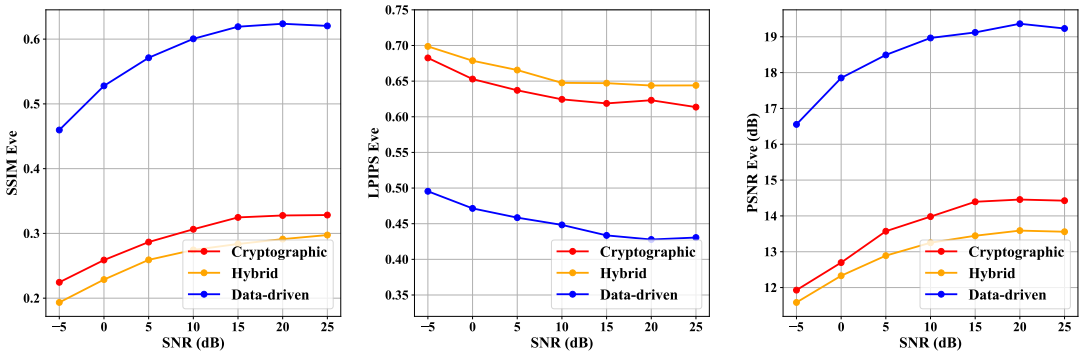

FIGURE 3.5: Eve's reconstruction quality across different protection methods on the SCUT-FBP5500 dataset.

maintaining coarse semantic outlines, allowing partial inference of image content. By contrast, the proposed Hybrid approach combines adversarial obfuscation with latent-space permutations, resulting in stronger visual degradation and preventing Eve from identifying any meaningful content.

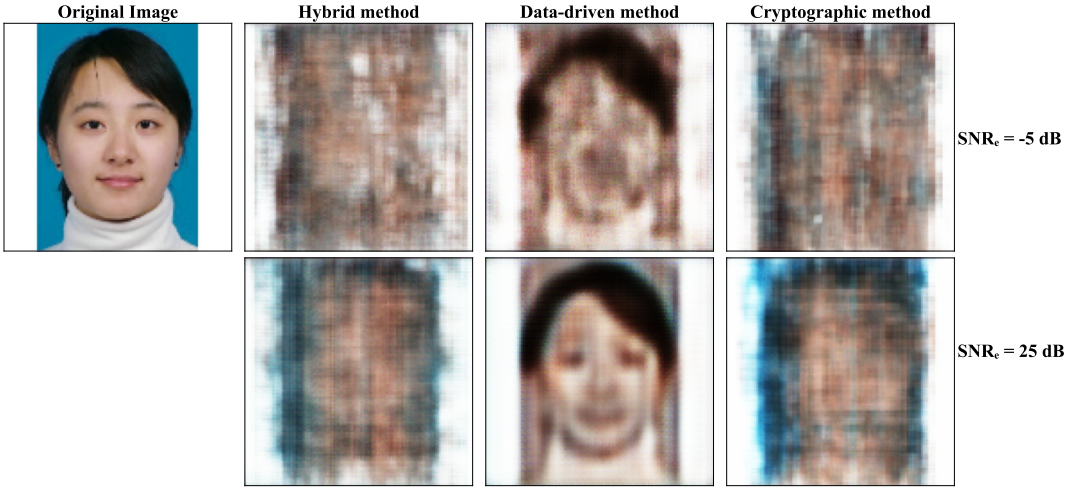

FIGURE 3.6: Visual comparison of Eve's reconstructed images under different protection methods for a randomly selected image from SCUT-FBP5500 dataset.

Table 3.3 further summarizes Eve's performance when both Bob and Eve are trained at SNR = 15 dB and evaluated at the same SNR under a Rayleigh slow fading channel, following the configuration in [9]. The results consistently show that even in this matched SNR scenario, the proposed Hybrid protection method effectively limits Eve's reconstruction capability.

Although these protection mechanisms enhance communication security, they may also impact Bob's reconstruction quality. Figure 3.7 presents Bob's performance under the various defense strategies, compared to the baseline JSCC model without protection. The cryptographic method yields reconstruction performance for Bob that aligns with JSCC, as its permutation and substitution operations are fully reversible; however, it requires a secure key exchange and reliable synchronization, which may not always be practical. On the other hand, the Hybrid approach offers a favorable

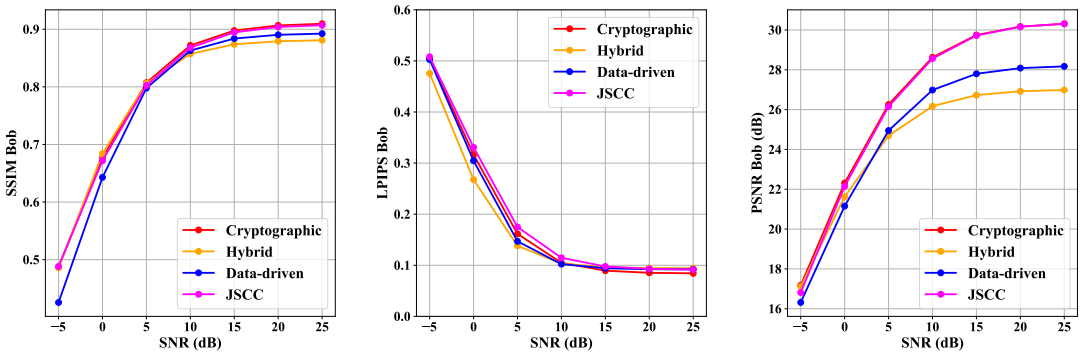

FIGURE 3.7: Impact of defense mechanisms on Bob's reconstruction performance across different SNRs evaluated on SCUT-FBP5500 dataset.

TABLE 3.3: Eve's reconstruction performance on the SCUT-FBP5500 dataset when $SNR_e = 15$ dB and $SNR_b = 15$ dB under a Rayleigh fading channel.

Method	**SSIM**	**LPIPS**	**PSNR**
None	0.60	0.37	19.46
Data-driven [46]	0.43	0.58	18.37
Cryptographic [37]	0.31	0.63	13.59
Hybrid	0.27	0.68	12.46

balance between reconstruction fidelity and privacy preservation, aligning well with the data-driven strategy while providing the strongest defense against eavesdropping. The relatively lower LPIPS values observed for the Hybrid and data-driven methods stem from their larger number of trainable parameters, which enable more accurate semantic reconstructions despite adversarial regularization.

3.3 KAP-DJSCC: A solution to MIEA

To counter MIEAs in JSCC, the early study in [37] utilizes permutation and substitution techniques applied directly to the latent representation. While this approach is resilient to both eavesdropping and channel noise, it requires twice the resources, as it sends two latent vectors instead of the original one. In addition, [118] applies steganography to create a semantically covert protection against MIEA. However, these mitigation methods mainly focus on data processing and architectural defense and do not integrate with telecommunication protocols. This limits their cross-layer adaptability and robustness in practical systems.

To the best of our knowledge, there are still some gaps in the literature that have not yet been addressed. Cryptographic methods [35], [37] can be resource-intensive, highlighting the need for simpler yet effective alternatives. Additionally, applying cryptography directly to the original image reduces its redundancy, which hinders compression using DNNs [35]. On the other hand, encrypting the latent space [35] also requires quantization, resulting in information loss. Moreover, data-driven models [44], [119] trade off full recovery and privacy by training the autoencoder with a

combined loss function. Since these methods are application-specific and the defense is not pluggable, the model must be retrained when full information is needed at the receiver's side. In addition, existing approaches in the literature often neglect authentication methods and focus solely on DNN architecture design against MIEA. This discussion highlights the need for a fast and lightweight algorithm that provides security of JSCC against MIEA by leveraging existing components in telecommunication protocols.

In this section, we propose a novel secret key assisted protection method for DJSCC, named as KAP-DJSCC to protect the semantic communication systems against MIEA. Our proposed method leverages the DH key exchange protocol to establish a shared secret key known only to the legitimate transmitter and receiver. Unlike conventional cryptographic approaches that rely on computationally intensive calculations and require quantization [35], the proposed scheme uses the shared key to generate a pseudo-random transformation matrix and applies a lightweight and reliable process to conceal the transmitted semantic information. Since the transform is fully reversible at the receiver side, the original data remains unaltered. This makes the method particularly suitable when accurate information recovery is required. Moreover, a novel MIEA, referred to as KMIEA has been proposed. Simulation results demonstrate that KAP-DJSCC effectively reduces the eavesdropper's ability to recover the original content through KMIEA and MIEA by lowering the similarity between the transmitted and source data.

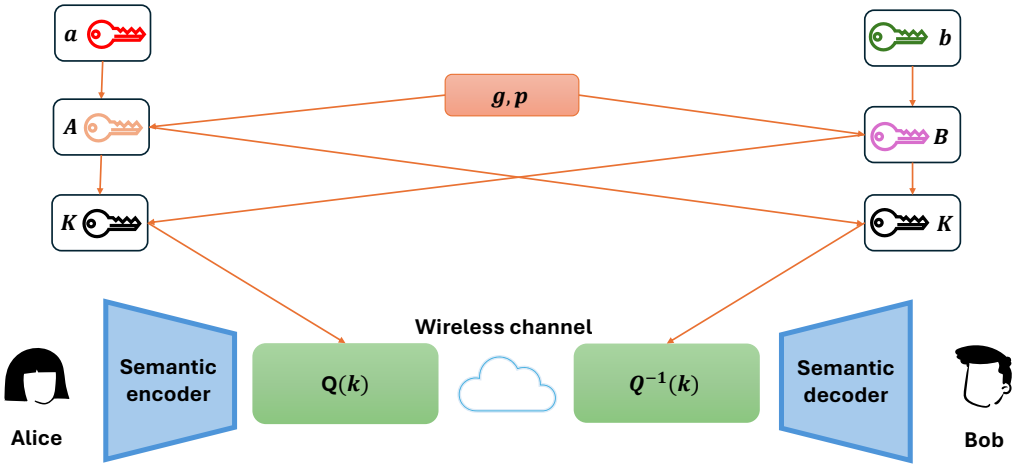

FIGURE 3.8: Defense model

3.3.1 Encoding

In JSCC, Alice employs an encoder $f : \mathbb{R}^N \rightarrow \mathbb{C}^c$ to extract semantic features from the input image $\mathbf{x}$ and compress them into a complex-valued latent vector $\mathbf{s}_{1\times c}$:

$$\mathbf{s} = f(\mathbf{x}; \phi), \tag{3.3}$$

where ϕ and c represent the trainable parameter set of the encoder and channel bandwidth according to [9], respectively.

3.3.2 Key Agreement via DH Protocol

Initially, Alice and Bob negotiate a shared secret key K using the DH protocol. They publicly agree on a large prime p and a generator g. Then, Alice selects a private key $a \in \mathbb{Z}_p$ and computes $A = g^a \bmod p$, while Bob selects a private key $b \in \mathbb{Z}_p$ and computes $B = g^b \bmod p$. After exchanging A and B, both compute the identical shared key:

$$K = B^a \bmod p = A^b \bmod p. \tag{3.4}$$

3.3.3 Key-based Transformation

The shared key K is used as a seed for a PCG64 pseudorandom generator [120]. Then, a random $c \times c$ matrix is generated and orthogonalized via QR decomposition. The orthogonal matrix $\mathbf{Q}$ which is suitable for geometric transformations, serves as a lightweight encryption tool for the latent vector. We refer to this process as $\Omega(k)$, which takes K and generates the invertible transformation matrix $\mathbf{Q} \in \mathbb{R}^{c \times c}$:

$$\mathbf{Q} = \Omega(K), \tag{3.5}$$

where $\mathbf{Q}^\top \mathbf{Q} = \mathbf{I}$. This matrix is then applied to the latent vector $\mathbf{s}$:

$$\mathbf{s}' = \mathbf{s}\mathbf{Q}^\top. \tag{3.6}$$

This reversible procedure preserves overall structure of the latent space but masks the latent representation from unauthorized reconstruction.

3.3.4 Transmission and Recovery

After the security module, the output $\mathbf{s}'$ is normalized to satisfy a given average power budget [9]. The normalized transformed latent vector is transmitted over an AWGN channel. The received signal at Bob's side is modeled as:

$$\hat{\mathbf{s}} = \mathbf{h} \circ \mathbf{s}'_n + \mathbf{n}, \tag{3.7}$$

where $\mathbf{h}$ is the channel gain vector, $\circ$ denotes element-wise multiplication, $\mathbf{s}'_n$ shows the latent vector after transformation and power normalization, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ is complex Gaussian noise.

Since Bob also possesses the shared key K and is aware of $\Omega(k)$, he reconstructs the orthogonal matrix $\mathbf{Q}$ and inverts the transformation to recover the original latent vector:

$$\tilde{\mathbf{s}} = \hat{\mathbf{s}}\mathbf{Q}. \tag{3.8}$$

3.3.5 Decoding

Bob then applies the decoder $f^{-1} : \mathbb{C}^c \to \mathbb{R}^N$ to reconstruct the original image from the latent representation:

$$\hat{\mathbf{x}} = f^{-1}(\tilde{\mathbf{s}}; \psi), \tag{3.9}$$

where ψ denotes the decoder's trainable parameters. The orthogonal nature of $\mathbf{Q}$ ensures that this transformation introduces no distortion, thereby preserving reconstruction quality.

In contrast, Eve without knowledge of the shared key cannot guess $\mathbf{Q}$ and hence cannot invert the transformation. This incapability significantly degrades Eve's ability to recover the semantic representation $\mathbf{s}$, providing enhanced confidentiality without altering the autoencoder's architecture or training procedure.

3.3.6 Training

In JSCC, the encoder and decoder are jointly trained to minimize the expected reconstruction loss over a training dataset $\mathcal{D}_{\text{train}} = \{\mathbf{x}^{(i)}\}_{i=1}^{M}$:

$$\mathcal{R}_M(\phi, \psi) = \mathbb{E}_{\mathbf{x}\sim\mathcal{D}_{\text{train}}}\left[\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}})\right], \tag{3.10}$$

where $\mathcal{L}(\cdot)$ is a chosen loss function. In this chapter we use MSE loss. Note that the encoder and decoder are trained jointly, whereas the key exchange and transformation steps are applied separately to the latent space as non-trainable operations. In essence, the protection module is designed to be pluggable. Figure 3.8 illustrates the DH-aided defense integrated with the JSCC.

3.3.7 KMIEA

In MIEA [37], [121], Eve queries the encoder illegally, receives the encoded data via its channel, and trains its *surrogate decoder*. We propose KMIEA, a *surrogate model training* [121] attack which trains its model after guessing K. In this attack, Eve leverages auxiliary information and illicitly accesses Alice's model predictions to train a surrogate model that approximates Bob's behavior. Since Eve does not know the true K, it uses a randomly chosen key K_e to simulate the defense mechanism. It is important to note that while Eve has no knowledge of Bob's decoder, it has full access to the encoder and knows the applied defense process in the transmitter's side. Since Eve does not know Bob's decoder or the reverse transform, it uses K_e to train its surrogate decoder on a transformed latent vector, enabling it to handle both the transformation and AWGN noise simultaneously. In fact, the model learns the mapping introduced by K_e. More specifically, KMIEA consists of the following steps:

- Query the encoder using an auxiliary training set D_{aux};
- Apply K_e to the queries using eq. (3.5) and (3.6).
- Receive the queries via the eavesdropping AWGN channel with its specific training noise $\mathbf{n}_e$;
- Use transformed noisy queries as inputs, and matching D_{aux} samples as targets, to form D_{adv} for inverse network training;
- Train the surrogate model that utilizes D_{adv} to approximate the inverse mapping.

For an image reconstruction attack in KMIEA, both the quality of D_{aux} and the design of the surrogate model are critical. Additionally, Eve's level of access to information about the encoder and the defense method significantly impacts the strength of the attack. All the steps employed in KMIEA and the training procedure are illustrated in Figure 3.9.

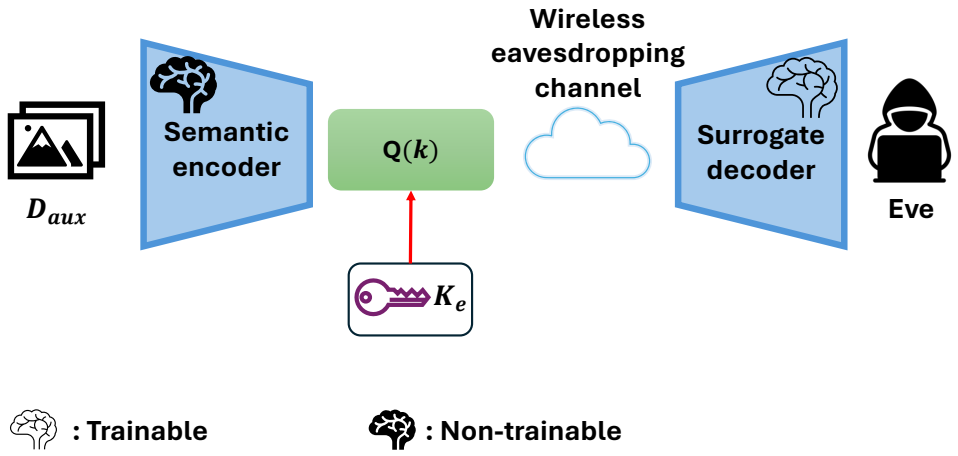

FIGURE 3.9: KMIEA training model

3.3.8 Simulation results

In our simulations, we evaluated performance using widely adopted image quality metrics, including pixel-wise measures such as PSNR and SSIM, as well as the perceptual similarity metric, LPIPS. The legitimate JSCC system is implemented based on the baseline architecture proposed in [9], while the architecture of Eve's surrogate decoder is presented in Table 3.4. All models are implemented in Python using the TensorFlow framework, and trained with a learning rate of 0.0001 for both Bob and Eve. The CIFAR-10 dataset is used, with 40000 images allocated for training Bob, 10000 images for training Eve's surrogate decoder, and 10000 images for testing. Eve and Bob are both trained at SNR of 20 dB.

TABLE 3.4: Eve's Decoder Architecture Summary

No.	Layer
1	Conv2DTranspose, 5×5, stride 1
2	PReLU
3	Conv2DTranspose, 5×5, stride 2
4	PReLU
5	Conv2DTranspose, 5×5, stride 2
6	Sigmoid

Figure 3.10 illustrates the reconstructed image quality evaluated using the aforementioned metrics across different channel SNRs. The curve labeled Bob, KAP-DJSCC represents the reconstruction quality at Bob's side when the proposed defense

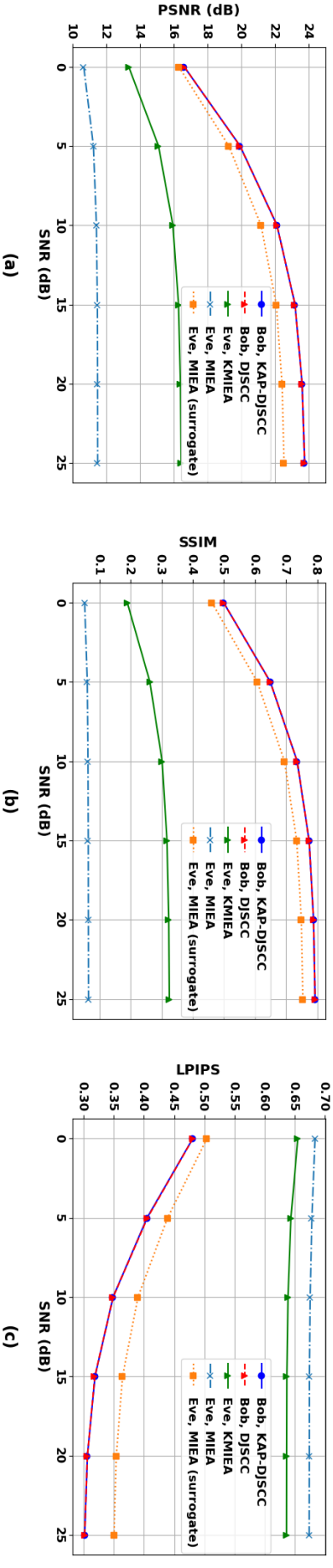

FIGURE 3.10: PSNR, SSIM, and LPIPS vs. SNR for Bob and Eve under different schemes. The proposed KAP-DJSCC method preserves high reconstruction quality for Bob while effectively degrading Eve's performance under both KMIEA and MIEA attacks.

mechanism is applied, while Bob, DJSCC shows the performance of the standard JSCC system without any protection. The curves Eve,KMIEA and Eve,MIEA depict Eve's reconstruction performance when KAP-DJSCC is applied under KMIEA and MIEA [37], respectively. Furthermore, Eve,MIEA (Surrogate) curve provides the comparison with the benchmark JSCC under MIEA. The results demonstrate that the proposed KAP-DJSCC method achieves image reconstruction quality nearly identical to the baseline JSCC in all subfigures. This confirms that the defense mechanism introduces no recognizable information loss or decoding failure, thanks to utilizing reversible transformation applied to the transmitted latent space. Moreover, Eve with MIEA can achieve results closely converging to Bob when KAP-DJSCC is not implemented. In contrast, the defense substantially degrades Eve's reconstruction quality. While key guessing (K_e) and training a surrogate decoder based on it in KMIEA improves Eve's ability, her performance remains significantly inferior to Bob. Since the secret key is exchanged via the DH protocol, successful guessing is highly improbable, ensuring robust protection against MIEAs.

In subfigures (a) and (b), the trends of the curves are similar, as both PSNR and SSIM are pixel-level quality metrics. Consistent with typical JSCC schemes [9], the proposed method and both attack models exhibit saturation at SNRs higher than the training SNR. In subfigure (a), the maximum degradation in Eve's reconstruction quality under KMIEA occurs at 25 dB, amounting to a 30.1 % reduction relative to the no-attack case, while under MIEA it amounts to a 51.7 % reduction. A similar trend is observed in subfigure (b), where decoding failure measured by SSIM peaks at 62.5% under KMIEA and 92.0% under MIEA. Unlike PSNR and SSIM, the lower values in LPIPS indicate better perceptual similarity. The defense method proves highly effective in this regard, substantially reducing perceptual reconstruction quality at Eve's side with maximum differences of 52.7% and 55.3% under KMIEA and MIEA, respectively. Although only the maximum differences are reported here, the performance gap between Bob and Eve remains consistently significant across all SNRs, demonstrating the robustness and effectiveness of the proposed KAP-DJSCC scheme.

Figure 3.11 illustrates qualitative reconstruction results at SNR levels of 20 dB and 0 dB for the proposed scheme and both attack models. At SNR = 20 dB, Bob with KAP-DJSCC recover the image with high fidelity compared to JSCC. The results confirm that the proposed DH-aided latent space transformation preserves reconstruction quality for Bob, maintains image texture, and accurately recovers the semantic content. In contrast, Eve's reconstructions under both attack models exhibit severe quality degradation when KAP-DJSCC is applied. KMIEA achieves relatively better values and reveals scrambled textures and color regions compared to MIEA, but the image still remains difficult to interpret. This proves the advantage of using a secret key exchange. Under MIEA, only noisy and unrecognizable outputs are obtained. In very low SNR regime (0 dB), Bob's reconstructions degrade slightly when KAP-DJSCC is applied while the semantic content remains clearly visible. Notably, Eve under MIEA produces no meaningful reconstruction, confirming that KAP-DJSCC effectively obscures the latent representations and protects the semantic contents against eavesdropping.

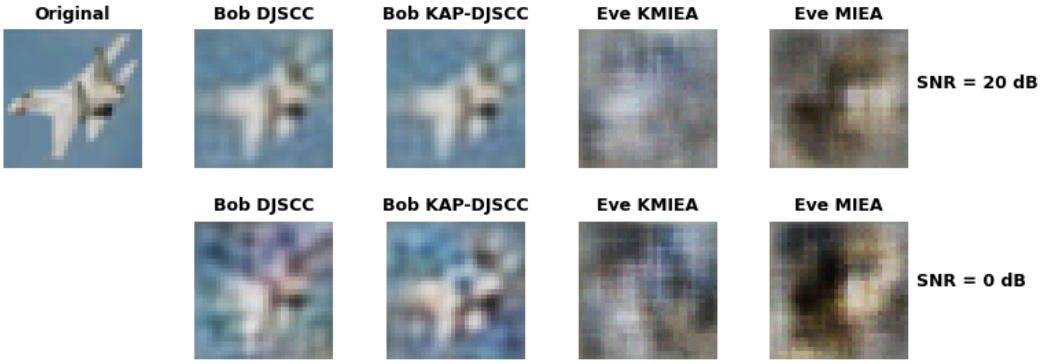

FIGURE 3.11: Reconstructed images at SNR = 20 dB (top) and 0 dB (bottom). Shown are outputs from Bob (with/without KAP-DJSCC) and Eve (KMIEA and MIEA) under KAP-DJSCC.

3.4 Open Challenges

Despite recent advancements in privacy-preserving techniques within SemCom, several critical challenges remain unresolved. This section outlines key open issues in semantic privacy protection and suggests promising directions for future research.

- **Application-aware architecture design:** In semantic and task-oriented communication, the focus shifts from accurate bit-level transmission to meaning extraction and task fulfillment. Consequently, the communication objectives, and thus the associated privacy threats, vary significantly across applications. This variability affects the methods for information extraction, transmission, reconstruction, and protection. In particular, PLS techniques must be tailored to the specific characteristics of each use case. Notable open challenges include secure semantic transmission in satellite imagery, integrated sensing and communication, safeguarding task outputs from eavesdroppers, and ensuring privacy in virtual transportation networks. Each of these scenarios requires a reevaluation of both task performance metrics and privacy-preserving strategies.

- **Trade-off between semantic security and network efficiency:** SemCom inherently supports privacy by transmitting only task-relevant information, reconstructed using the receiver's KB. However, this benefit comes at the cost of increased computational complexity [52]. This is particularly problematic in large-scale or resource-constrained deployments, where energy efficiency and sustainability are critical. In remote or battery-operated networks, minimizing power consumption without sacrificing security or task performance is a major challenge. Moreover, additional security layers may degrade inference accuracy or reconstruction quality [42], [122]. Thus, future research should focus on designing adaptive frameworks, loss functions, and training strategies that balance semantic security with computational and resource constraints.

- **Security-aware semantic performance metrics:** Traditional communication metrics such as Symbol Error Rate (SER) and Bit Error Rate (BER) are insufficient for evaluating SemCom systems, as they do not reflect the semantic integrity of transmitted information. Instead, task- and content-aware metrics are required. For

instance, semantic rate has been proposed for text transmission [77], yet it lacks provisions for evaluating security or secrecy. A comprehensive suite of metrics is needed—particularly for rich media such as video—to jointly assess semantic accuracy and information leakage. The survey in [123] provides a starting point, but more rigorous formulations are necessary to incorporate both semantic fidelity and privacy guarantees in future systems.

- **Active eavesdroppers and semantic adversarial attacks:** Unlike passive threats, active eavesdroppers can launch targeted semantic attacks by intercepting and modifying data in ways that maintain surface-level semantic similarity while altering core meaning [124]. Such attacks can mislead task execution at the receiver (e.g., decision-making at Bob). For instance, [125] examines active interference from a warden attempting to capture confidential UAV-ground communication. Adversarial attacks, as discussed in [111], further increase the risk by enabling simultaneous data theft and manipulation. Despite their severity, these threats remain underexplored in the SemCom literature. Future work should develop robust countermeasures against such intelligent and adaptive attackers.

- **Adaptability to unknown or mobile eavesdroppers:** A common limitation in current approaches, especially in PLS and data-driven methods, is the reliance on prior knowledge of the eavesdropper's location or channel conditions. For example, authors in [46] assume that the eavesdropper's channel is always weaker than the main link. However, mobile or adaptive eavesdroppers can easily violate these assumptions, leading to degraded privacy performance. To enhance the real-world applicability of secure SemCom, future techniques must be designed to operate effectively without relying on such assumptions, adapting dynamically to unknown and evolving threat environments.

- **Extension to large-model agent networks:** With the rapid advancement of general-purpose intelligent agents built upon large models, SemCom is expected to play a crucial role in both inter-agent and intra-agent communications. These models operate in a semantic-aware, task-oriented manner, where agents first interpret meaning and then communicate only the information required for the task. Adapting to task demands is a core property of large-model agent networks. However, this evolution also exposes new security vulnerabilities, particularly the risk of active or passive semantic eavesdropping on the communication links among these agents. As these networks predominantly rely on natural language processing and high-level semantic exchanges, conventional protection techniques often fail to satisfy their stringent security requirements for safeguarding semantic information. Therefore, a promising direction for future research lies in the development of advanced and adaptive security mechanisms specifically designed to counter eavesdroppers in large-model agent networks [126].

3.5 Conclusion

Unlike traditional communication systems that compress, transmit, and reconstruct data at the bit level, SemCom prioritizes meaning and task relevance throughout the

transmission process. By leveraging AI to extract and transmit semantic features wirelessly, SemCom introduces new challenges that arise from both AI and wireless communication in the semantic domain. A critical concern in this domain is eavesdropping, in which an unauthorized receiver accesses the transmitted stream and attempts to infer its meaning, resulting in privacy leakage. This chapter provided an analysis on the capability of a MIEA network based on the eavesdropper's access to the encoder's training parameters. The effectiveness of data-driven protection, permutation and substitution techniques, and their combination in securing JSCC against a representative adversarial architecture was evaluated. Results indicate that the combined approach significantly degrades eavesdropping performance in terms of PSNR and SSIM, while the cryptographic method imposes less complexity.

Moreover, this section introduced KAP-DJSCC, a lightweight and reversible defense framework that safeguards semantic image transmission via JSCC against MIEA. Using a DH-based key exchange protocol, the proposed method applies a key-assisted orthogonal transformation to the transmitted latent space, which enhances security without altering the original meaning or requiring retraining. To assess the efficiency, we introduced a new attack, KMIEA, which performs key guessing and surrogate model training. The simulations using PSNR, SSIM, and LPIPS metrics demonstrate that KAP-DJSCC preserves high reconstruction quality at the legitimate receiver while significantly degrading the eavesdropper's performance across all channel SNRs. Visual results also confirm the effectiveness of KAP-DJSCC in protecting semantic content.

Finally, the section outlines critical research gaps and open challenges, offering guidance for future research in secure semantic communication. For future work, a more advanced DH method such as Elliptic-Curve DH is recommended. The basic DH, although efficient, remains vulnerable to man-in-the-middle attacks, which could also serve as an entry point for multidomain attacks.

Chapter 4

Risk-Based Adaptive Security Framework for 6G Edge

4.1 Introduction

While 6G in mobile communication is not yet standardized, it promises numerous new features and benefits, such as ultra-high data rates (up to 1 Tbps), ultra-low latency, high energy efficiency, ubiquitous connectivity, and the pervasive integration of AI [127]. These networks aim to provide ubiquitous connectivity across ground, aerial, maritime, and underwater environments. This enables public services such as remote learning and telemedicine to operate seamlessly across geographical boundaries [128].

Although 6G offers numerous promising capabilities, it introduces new security and privacy challenges. As new services and technologies are adopted, the threat landscape evolves accordingly which can significantly impact the performance of these networks [129], [130].For instance, zero-day attacks, AI- and ML-based attacks, quantum threats, and physical-layer attacks represent new types of threats that can compromise 6G networks [130]. To address emerging challenges, new security strategies must be developed to safeguard the network. The solutions utilized for 5G are no longer suitable for 6G due to the greater diversity in device capabilities, service requirements, energy constraints, attack surfaces, and other time-varying attributes. In fact, 5G security solutions often rely on a one-size-fits-all strategy across devices and base stations configured with the same security settings which is not applicable to 6G networks [131].

Like other cellular networks, the cybersecurity risks in 6G result in the loss of confidentiality, integrity, or availability of information or control systems. This potentially leads to adverse impacts on organizational operations, including reputation, assets, individuals, other organizations, and even national security [132]. Thus, recognizing the risks associated with cyber incidents is essential for network operators and consists of the first step toward implementing appropriate security measures for the protection of personal data, according to General Data Protection Regulation (GDPR) framework [133].

Furthermore, 6G platforms face limitations such as constrained bandwidth and limited computational, energy, and financial resources. In such highly constrained environments, deploying edge servers and offloading computations to the cloud can help reduce computational load and ensure high data rates in densely connected areas. In fact, edge computing, edge caching, and edge intelligence are expected to be widely

adopted to extend cloud capabilities closer to the assets [134]. To satisfy the security needs in this highly distributed environment with edge computing, cloud computing, and assets, a new framework called Secure Access Service Edge (SASE) has been proposed in [135]. The goal of SASE is to combine comprehensive Wide Area Network (WAN) capabilities with network security functions [136]. Since SASE can be deployed across a wide range of devices due to its scalability and flexibility, it would be highly suitable for securing 6G networks. It builds upon the concept of Software-defined Wide Area Network (SDWAN) by integrating security into the network architecture. Moreover, it unifies various network functions, such as Virtual Private Network (VPN), with advanced security services including Zero-Trust Architecture (ZTA), Firewall as a Service (FaaS), and Secure Web Gateway (SWG). With the increasing adoption of edge computing, network operators are showing a growing interest in SASE solutions. Moreover, the concept of *Security-as-a-Service (SaaS)* proposed by SASE offers a cloud-based service to manage cybersecurity tools and their costs. SASE provides compatible, real-time defense against threats and the possibility to secure data within all environments [137]. As an example in the context of 6G, implementing a ZTA can effectively enhance data security [138] for the system level such as UAVs, demonstrating the potential of SASE in securing air–ground integrated networks.

Moreover, integrating space, aerial, and terrestrial segments in future communication networks introduces new security challenges. Small-satellite systems create an intermediate layer between terrestrial infrastructures and traditional satellite systems, providing relay-assisted connectivity, coverage extension, and potentially new security capabilities [139]. As noted in [139], aerospace scenarios further exacerbate security requirements; consequently, existing security protocols, cryptographic mechanisms, and physical-layer security techniques must be revisited and adapted to the unique characteristics of integrated space–air–ground networks.

Building on the earlier discussion about device limitations in 6G, it becomes clear that security measures must be adaptive to account for constrained resources and associated costs. Countermeasures should be deployed rapidly, while maintaining a balance between competing requirements and limited network resources. In other words, the optimal set of security features should be selected in real time based on current cybersecurity risk exposure and the available budget. These considerations highlight the importance of an adaptive security framework to manage the trade-off between security and cost. In particular, cyber risk assessment must be performed dynamically and in real time to support this balance. The risk assessment frameworks must account for assets, vulnerabilities, and threats to effectively quantify and measure the risk. The model's cost should also be minimized, while maintaining flexibility to accommodate both small- and large-scale networks [140]. In particular, given the resource constraints in NTN and across space–air–ground devices, it is promising to develop resource-aware security strategies and leverage edge and cloud computing [141].

To the best of our knowledge, and based on the preceding discussion, there is a need for an adaptive framework that builds on existing risk assessment methods and extends them to manage constrained resources in 6G. This framework must also be comprehensive enough to guide the selection of optimal countermeasures. Several frameworks in the literature focus on 5G-related security challenges [129], [134],

[142], or address risk assessment exclusively for specific types of networks like IoT [143], rather than for 6G systems [140], [144]. In contrast, some recent studies acknowledge 6G and resource limitations while they still lack comprehensive risk assessment mechanisms [131]. Additionally, other works propose frameworks for 6G [145], [146], [147], [148], but they overlook the impact of such constraints. Moreover, existing studies such as [149], [150] concentrate on countering specific attacks in 6G networks; nevertheless, a unified framework for the systematic selection of countermeasures is still missing.

To estimate cybersecurity risk, Information Security Forum (ISF) proposed the Information Risk Assessment Methodology 2 (IRAM2) framework, an analytical model that offers a practical, business-oriented approach for end-to-end risk management. It has a business perspective while it is uniformly understandable. IRAM2 allows for identifying the significant risks and help managers to allocate resources. Moreover, it can engage stakeholders. According to IRAM2, assessing and addressing information risk involves six stages: scoping, impact assessment, threat profiling, vulnerability assessment, risk evaluation, and risk treatment [151]. Building on the six phase process suggested by ISF, a novel formulation of the risk function is proposed which evaluates the cybersecurity risk within the IRAM2 framework. Subsequently, the risk treatment is formulated as an optimization problem, which is solved using various methods, including linear programming, Genetic Algorithm (GA), and innovative heuristics. In particular, the main strength of the proposed framework lies in its *adaptive* nature via a *real-time* cyber risk estimation. Accordingly, the main contributions of this chapter are:

- A new secure risk-based framework for 6G based on SASE has been proposed, which enhances security by assessing the risk and leveraging the core capabilities of SASE for defense. This cloud-native approach enables the system to respond to dynamic 6G environments, particularly when operating with resource-limited devices, for instance, in scenarios involving the integration of aerial and ground platforms.

- A threat modeling and risk assessment process following the IRAM2 guidelines is proposed, which is chosen for its flexibility and scalability. This approach is suitable for scenarios of all sizes and can be integrated with other risk management tools. In this context, a novel assessment methodology has been proposed that leverages both measured and estimated attack probabilities, incorporates impact assessment, and evaluates the resulting risk.

- To determine and select the optimal set of countermeasures against the risk evaluated within IRAM2 in real-time, several heuristics and optimization methodologies have been proposed. This chapter aims to identify the countermeasure set that minimizes the maximum value of the risk vector across all threats, subject to the system's available budget. The computations are designed to be implemented in the edge and the SASE platform provides the adaptive security for the system.

The remainder of this chapter is organized as follows. Section 4.2 provides a brief summary of related works. Section 4.3 presents our model. In Section 4.4, simulation results are presented and finally Section 4.5 concludes the chapter.

4.2 Related works

This section presents our literature review on existing security and risk assessment frameworks, with a focus on their application in 5G and 6G edge networks, as well as the integration of SASE into communication systems. To support 6G applications, edge computing, edge caching, and edge intelligence are widely adopted to store and process local data, as summarized in the survey by [129]. Edge computing and caching is already adopted in 5G networks in order to have mini-clouds closer to the User Equipment (UE) [142]. Although these three techniques can assist in reducing response time and bandwidth usage, the threats in the network yet exists while introducing AI adds new threats. Additionally, the limited resources prevent the edge devices and assets to implement powerful protection methods [129], [152]. Although edge computing has greatly benefited 5G, it also introduces specific security challenges. Due to its heterogeneous nature, reliance on wireless systems, and use in applications that require high processing power, it is vulnerable to intrusions, unauthorized access, and a wide range of security attacks [134]. On the other hand, these threats impose a risk on the system which have been analyzed in [145]. As a result, relying on security frameworks that select the best countermeasures dynamically based on the new constraints in 6G, while leveraging the cloud for defense, is beneficial for future networks.

As mentioned in the previous section, the *one-size-fits-all* strategy is easy to implement but lacks generalizability for 6G systems, as 6G devices are more diverse and deployed across both accessible and remote areas, requiring security solutions to be appropriately tailored. Moreover, the evolving nature of services, exposure to various attacks, and dynamic system behaviors in 6G necessitate a security strategy that is both adaptable and specifically designed for the unique demands of 6G [131]. To mitigate attacks, researchers have explored threat and countermeasures modeling techniques and conducted comprehensive risk assessments to identify potential vulnerabilities in 6G networks [146], [147]. However, most of the studies focus on optimizing security [147], applying defenses like encryption [147], designing strategies for a single attack type [149], or edge intelligence-based authentication [150] without conducting thorough risk assessments accordingly.

The topic of adaptive and dynamic security in 6G has been investigated in [131] with a focus on energy efficiency. The upcoming security threats in 6G from the dynamics, heterogeneity, and complexity point of view have been analyzed and an optimization framework is proposed to tackle these issues. However, while the work offers energy-aware adaptive solutions, it overlooks the consideration of attack probabilities and lacks a comprehensive risk analysis.

A dynamic risk assessment framework has been proposed in [145] for 5G which utilizes a closed-loop feedback to monitor the threats, and analyze them. Since 5G-IoT has a dynamic characteristic, the work in [145] suggests that the security services must be adaptive. However, the quantitative analysis of the risk and impact of the threats on the network is not analyzed and the proposed framework cannot be generalized to 6G due to the expanded nature of 6G. In contrast, [146] designs a framework specially tailored for 6G, which provides a human-centric model for the security purposes. By leveraging an AI-based smart security layer, they immediately detect the

threats and correct the vulnerabilities. Nevertheless, the authors in [146] do not consider the possible limitations in 6G like budget and bandwidth. In addition, [147] utilizes the network-slicing architecture to protect brand design data and protect privacy using encryption, anomaly detection, and securing network slices. Moreover, [149] investigates DDoS attacks in service function chains within 6G networks. The authors propose a blockchain-based algorithm to validate virtualized network functions and employ reinforcement learning to solve the corresponding attack–defense game. [150] presents a collaborative edge-intelligence authentication scheme for attacker identification and localization. In this approach, edge nodes support the service provider in user authentication, and an attacker localization algorithm is further proposed. This method improves the security by utilizing multi-dimensional security information and multiple cooperating devices.

In the era of bridging the gap between qualitative and quantitative analysis in risk assessment, the reference [153] proposes a probabilistic method Threat Informed cyber resilience index for evaluating the defense effectiveness against cyber attacks. The partially observable Markov decision processes have been employed to model attacker's behavior. Although this method makes the owners to make data-driven decisions, it does not provide a general framework for risk assessment and taking decisions according to the limitations. The authors of [143] investigated Advanced Persistent Threats (APTs) and their impact on IoT networks. They examined the risk of APT propagation among IoT devices and modeled attacker behaviors. To analyze the interdependence between cyber insurance providers, IoT defenders, and APT attackers, they employed a game-theoretic approach. Solving the optimization problem yields the best insurance contract. Moreover, [140] introduces a new decision-making framework that evaluates risk probability in line with ISO/IEC 27005:2018 [154] for enterprise networks. While it automates the risk assessment process, it overlooks the efficient selection of countermeasures and is tailored specifically for enterprise environments rather than 6G systems. The authors in [144] prioritize security countermeasures by ranking vulnerabilities, exploits, and privileges using a tripartite graph model. The framework enables indirect resource allocation through informed prioritization; however, it does not explicitly model resource constraints or incorporate them into a real-time optimization problem.

In the context of applying SASE to networks, authors in [148] have proposed an architecture that integrates network security and services based on SASE principles within 5G LANs. This design enhances the architecture's efficiency, manageability, global accessibility, resilience, connectivity, and adaptability. Additionally, the use of SD-WAN further improves efficiency by offloading computing tasks to the cloud. [148] presents the capabilities of SASE specifically for 5G LANs. Nevertheless, it does not consider security risk factors or how to tailor them to 6G. Moreover, [138] proposes a programmable ZTA to achieve dynamic and fine-grained security control. Authors in [143] have investigated APT in IoT networks. Moreover, [138] introduces a programmable ZTA framework, called SYSFLOW, designed to provide dynamic and fine-grained security control. SYSFLOW enables risk awareness, microsegmentation, and other essential security functions, extending ZTA capabilities beyond the network layer to encompass system-level protection. The framework leverages a logical controller that provides a policy decision point to regulate access to

system resources. As noted in [138], the focus is solely on ZTA, whereas SASE encompasses ZTA along with additional security services.

To the best of our knowledge, there is a lack of a general framework that models threat probability and impact, assesses vulnerabilities, estimates risk based on observations and predicted attack probabilities, and identifies the optimal set of defense countermeasures while accounting for potential budget constraints. In this chapter, we explore adaptive security in 6G networks from a risk-based perspective, aiming to balance the risk–cost trade-off across diverse scenarios.

4.3 Proposed Framework

As previously discussed, there is a need for a comprehensive framework that investigates cyber risk under constrained resources and selects the optimal set of countermeasures to mitigate cyber attacks. To address this problem, following the guidelines of IRAM2 is beneficial for both qualitative and quantitative risk evaluation. The proposed framework adheres to the IRAM2 steps, namely: impact assessment, threat profiling, vulnerability assessment, risk evaluation, and risk treatment, under these assumptions:

TABLE 4.1: Variables of the system model.

Notation	**Description**
N	Number of threats.
K_T	Number of terrestrial assets.
K_{NT}	Number of non-terrestrial assets.
$K = K_T + K_{NT}$	Number of assets.
M	Number of countermeasures.
B	Vector of budget, including financial budget, RAM, CPU, and energy.
$T = \{t_1, \ldots, t_N\}$	Set of threats.
$C = \{c_1, \ldots, c_M\}$	Set of countermeasures.
$A_T = \{a_1^{(T)}, a_2^{(T)}, \ldots, a_{K_T}^{(T)}\}$	Set of terrestrial assets.
$A_{NT} = \{a_1^{(NT)}, a_2^{(NT)}, \ldots, a_{K_{NT}}^{(NT)}\}$	Set of non-terrestrial assets.
$A = A_T \cup A_{NT}$	Set of assets.
$I = \{i_1, \ldots, i_K\}$	Impact vector.
$S \subseteq C$	Strategy, i.e., the subset of used countermeasures.
$\Phi \in \mathbb{R}^{M \times B}$	Set of countermeasures' costs.
p_n	Probability of attack for the n-th threat.
$p_n^{(E)}$	Probability of attack based on the environment.
$p_n^{(M)}$	Probability of attack based on internal detection.
q_n	Probability of success for the n-th threat.
$\zeta_{K \times N}$	Asset exposure matrix.
$E_{N \times M}$	Countermeasures effectiveness matrix.
$\Theta \in \mathbb{R}^{N \times M}$	Threat strength matrix.
R	Residual strength matrix.
$\boldsymbol{d}$	Damage vector.

4.3.1 Scoping

The first phase of IRAM2, known as the scoping phase, aims to establish a clear, business-driven perspective on the risks. In this step, a profile of the environment is developed. We consider a scenario in which risk is assessed for a 6G network that integrates UAVs with a ground-based edge server. The edge server performs the assessment within IRAM2, while it incorporates SASE for implementing the selected countermeasures.

4.3.2 Impact Assessment

In phase B of IRAM2, the potential business impact is assessed if critical aspects of information assets or systems are compromised. To achieve this, an assessment method is proposed, which utilizes the damage vector and asset exposure (ζ) matrix to assess the impact of cyber threats on business operations. In the event that asset a_k is successfully compromised by a specific threat, the resulting impact can be categorized into three dimensions: economic, reputational, and legal. Accordingly, for each asset $a_k \in A$, the associated impact can be represented by a damage vector $\boldsymbol{d}_k$ as follows:

$$\boldsymbol{d}_k = [d_k^{\mathrm{E}}\ d_k^{\mathrm{R}}\ d_k^{\mathrm{L}}]^T, \tag{4.1}$$

where d_k^{E}, d_k^{R}, and d_k^{L} denote the economic, reputational, and legal damages for asset a_k, respectively. Each of these damage components has three severity levels: high, medium, or low.

$$d_k^{\mathrm{E}},\ d_k^{\mathrm{R}},\ d_k^{\mathrm{L}} \in \begin{cases} 3, & \text{High} \\ 2, & \text{Medium} \\ 1, & \text{Low} \end{cases} \tag{4.2}$$

Accordingly, the damage vectors corresponding to all assets $k \in [1, K]$ including terrestrial and non-terrestrial ones are grouped by impact category:

$$\begin{aligned} \boldsymbol{d}^{\mathrm{E}} &= [d_1^{\mathrm{E}}, \ldots, d_{K_T}^{\mathrm{E}},\ d_{K_T+1}^{\mathrm{E}}, \ldots, d_K^{\mathrm{E}}] \\ \boldsymbol{d}^{\mathrm{R}} &= [d_1^{\mathrm{R}}, \ldots, d_{K_T}^{\mathrm{R}},\ d_{K_T+1}^{\mathrm{R}}, \ldots, d_K^{\mathrm{R}}] \\ \boldsymbol{d}^{\mathrm{L}} &= [d_1^{\mathrm{L}}, \ldots, d_{K_T}^{\mathrm{L}},\ d_{K_T+1}^{\mathrm{L}}, \ldots, d_K^{\mathrm{L}}] \end{aligned} \tag{4.3}$$

Since each asset is vulnerable to a specific subset of threats $\tau \subseteq T$, the binary matrix $\zeta_{K\times N}$ is defined as follows:

$$\zeta_{kn} = \begin{cases} 1, & \text{if asset } k \text{ is vulnerable to } n\text{-th threat} \\ 0, & \text{otherwise} \end{cases} \tag{4.4}$$

After calculating the economic, reputational, and legal damages, the respective impact can be computed using the matrix ζ. The economic, reputational, and legal impact matrices, denoted as $\mathbf{I}^{\mathrm{E}}, \mathbf{I}^{\mathrm{R}}$, and $\mathbf{I}^{\mathrm{L}}$, are obtained by performing element-wise multiplication between the corresponding damage vector and each column of ζ, such that:

$$\begin{aligned} \mathbf{I}^{\mathrm{E}}(:, n) &= \boldsymbol{d}^{\mathrm{E}} \circ \zeta(:, n), \quad \forall n = 1, \ldots, N \\ \mathbf{I}^{\mathrm{R}}(:, n) &= \boldsymbol{d}^{\mathrm{R}} \circ \zeta(:, n), \quad \forall n = 1, \ldots, N \\ \mathbf{I}^{\mathrm{L}}(:, n) &= \boldsymbol{d}^{\mathrm{L}} \circ \zeta(:, n), \quad \forall n = 1, \ldots, N \end{aligned} \tag{4.5}$$

Then, the total impact $\boldsymbol{I} \in \mathbb{R}^{1\times N}$ is assessed by first identifying the maximum impact in each category across all assets for each threat, followed by computing their weighted sum, where $\gamma_1 + \gamma_2 + \gamma_3 = 1$:

$$\boldsymbol{I} = \gamma_1 \max_{k=1,\dots,K} \mathbf{I}^{\mathrm{E}}(k,:) + \gamma_2 \max_{k=1,\dots,K} \mathbf{I}^{\mathrm{R}}(k,:) + \gamma_3 \max_{k=1,\dots,K} \mathbf{I}^{\mathrm{L}}(k,:) \tag{4.6}$$

In $\boldsymbol{I}_{1\times N}$, each element i_n represents the total impact on the system in the n-threat succeed.

4.3.3 Threat profiling

To model the threats, we begin by leveraging an approximate probability of their occurrence. As previously noted, accurately determining p_n^R is infeasible. Therefore, we propose a novel mathematical framework to estimate the probability of an attack based on both an a priori probability and an observed measure.

Calculating Probability of attack

Each threat n occurs with a certain probability in real-world p_n^R, which depends on the attackers' intent and is difficult to predict. This probability is modeled as follows:

- An a-priori estimated probability p_n^E is considered, assumed to follow a uniform distribution.
- The edge server detects incoming threats, monitors them over a time frame T_f, and estimates the real-time probability p_n^M. Within T_f, the measurement unit records whether an attack occurred in each timeslot (assigning a value of 0 or 1), and then computes a weighted average over the elapsed timeslots to obtain p_n^M.

Finally, the approximated probability of n-th threat (p_n) is calculated as:

$$p_n = (1-\alpha)p_n^E + \alpha p_n^M, \quad 0 < \alpha < 1 \tag{4.7}$$

where α serves as a tunable parameter balancing the statistical estimated probabilities and real-time detections.

4.3.4 Vulnerability assessment (Calculating the probability of success)

Let p_n denote the approximated probability that the n-th threat occurs. The probability that the n-th threat successfully carries out an attack, denoted by q_n, depends on the inherent strength of the threat and the remaining attack capability after the implementation of countermeasures.

Let's consider $\boldsymbol{C} = \{c_1, .., c_M\}$ the set of all available countermeasures. Each countermeasure m has a certain level of effectiveness against the n-th threat, defined as $0 < e_{nm} < 1$. In other words, this is the probability that the m-th countermeasure can protect the system against the n-th threat. Thus, We can define a countermeasures

effectiveness matrix as $\boldsymbol{E} = [e_{nm}]$, with $n = 1, \ldots, N$ and $m = 1, \ldots, M$.

Since each countermeasure incurs an associated cost, the edge server selects an optimal subset of countermeasures based on the available budget constraint. The binary set of activated countermeasures defines the strategy $\boldsymbol{s}$, a $1 \times M$ vector whose elements indicate whether each countermeasure is active (1) or inactive (0).

$$s_m = \begin{cases} 1, & \text{if countermeasure } m \text{ is activated} \\ 0, & \text{otherwise} \end{cases} \tag{4.8}$$

for each $m = 1, \ldots, M$. After selecting the strategy s, the *control strength* associated to that s is computed as:

$$\boldsymbol{cs}(n, :) = \boldsymbol{E}(n, :) \odot \boldsymbol{s}, \quad \forall n = 1, \ldots, N \tag{4.9}$$

When a strategy is implemented to protect the system, each threat n has a *residual strength*, that is the residual probability of success for the threat, conditional on the threat occurring. In particular, $r_{nm} = 1 - cs_{nm}$ is the residual strength of the n-th threat given that m-th countermeasure is implemented.

Considering that countermeasures can be assumed independent of each other, the *aggregated residual strength* for strategy s is defined as:

$$\boldsymbol{ars}(n) = \prod_{m=1}^{M} r_{nm}, \quad \forall n = 1, \ldots, N \tag{4.10}$$

.

Finally, the success probability for the threat n under strategy s is equal to the approximated threat probability multiplied by the aggregated residual strength:

$$q_n^s = p_n \, \boldsymbol{ars}(n) \tag{4.11}$$

where $q^s = [q_1^s, \ldots, q_N^s]$ denotes the vector of success probability for all threats when strategy vector s is applied.

4.3.5 Risk evaluation

According to the fifth stage of IRAM2, the risk function associated with a particular threat n, is defined in this subsection. To do this, a quantitative estimation of the damage the edge suffers if the threat succeeds is required. A new function for assessing the risk is proposed here, which leverages the Impact i_n and the probability of the threat success q_n.

$$\text{Risk}_n^s = i_n \, q_n^s. \tag{4.12}$$

4.3.6 Optimization Problem (Risk treatment)

In the final step, the risk function defined previously must be mitigated. The aim of the proposed framework is to find the optimal s that minimizes the maximum of the

$$\begin{aligned} &\underset{S}{\text{Min}} \quad \underset{n=1,\ldots,N}{\text{Max}} \quad I_n \cdot P_n \cdot \Theta_n \cdot \prod_{m=1}^{M} \left(1 - E_{n,m}.s_m\right), \quad \text{for } n = 1, \ldots, N \\ &\text{subject to} \quad s\Phi \leq \mathbf{B} \end{aligned} \tag{4.14}$$

risk vector among all possible threats. In fact, the risk treatment is defined as the following optimization problem:

$$\begin{aligned} &\underset{S}{\text{minimize}} \quad \underset{n=1,\ldots,N}{\text{maximize}} \quad \text{Risk}_n^s \\ &\text{subject to} \quad s\Phi \leq \mathbf{B} \end{aligned} \tag{4.13}$$

The parameters used in the optimization problem are defined and Table 4.1. More precisely, the risk treatment problem is defined as the problem shown in (4.13) which can be solved by exhaustively searching all possible states. This solution is the optimal way of solving the problem, but it is time-consuming and the delay in processing can lead to severe consequences and high impacts on the system. To find the best strategy faster, we propose solving (4.13) using simplified approximations to achieve fast and near-optimal results.

4.3.7 Proposed solution (MaxScoring algorithm)

In this subsection, we propose a novel heuristic algorithm that searches for a high-quality solution to the optimization problem. The algorithm jointly considers all key variables in the main formulation, including probabilities, effectiveness, and cost. Rather than solving the main problem directly, it iteratively explores candidate solutions and continues until the best result is obtained which is definitely faster than the full search.

Initially, the selected strategy S is set as a vector of zeros and it gets updated in each iteration. Consider the element-wise multiplication of I, P, and Θ, and let j denote the index of the countermeasure activated in the previous step. Then, $\nu_{1\times N}$ is defined as:

$$\nu_i = I_i \cdot P_i \cdot \Theta_i \cdot \prod_{j=1}^{M} \left(1 - E(i,j)\right), \quad \text{for } i = 1, \ldots, N$$

In addition, take $CS_{1\times M}$ defined as:

$$CS = \sum_{i=1}^{N} E(i,j) \cdot \nu(i)$$

In addition, let $\Phi_{n,1\times M}$ be the normalized form of Φ defined as:

$$\widehat{\Phi} = \Phi \oslash \left(\mathbf{1}_{M\times 1} \cdot \left(\mathbf{B} - \sum \Phi(j,:)\right)\right) \tag{4.15}$$

$$\Phi_n = \sum_{i=1}^{b} \widehat{\Phi}_{i,j} \tag{4.16}$$

$$CSC_j = \frac{CS_j}{\Phi_{n,j}} \tag{4.17}$$

$$j^* = \arg\max_j \, CSC_j \tag{4.18}$$

After calculating $\Phi_{n,1\times M}$, the division of the CS and Φ_n will be employed to initially introduce a parameter to find the most effective countermeasure with lower cost. After calculating the maximum CSC among the available countermeasures, the countermeasure j that maximizes CSC is selected in each iteration. This process continues until no further countermeasure can be selected within the budget $\mathbf{B}$.

4.3.8 Benchmarks

Linear estimation

Since the optimization problem in formula (4.13) is nonlinear, it is not mathematically tractable. Approximating the multiplication term with a linear summation reduces the problem's complexity, enabling faster convergence and making it possible to solve using Mixed Integer Linear Programming (MILP). Therefore, the linear estimation of the nonlinear term in (4.13) is presented as:

$$\prod_{j=1}^{M} (1 - E_{n,j} \cdot s_j) \approx 1 - \sum_{j=1}^{M} E_{n,j} \cdot s_j \tag{4.19}$$

After linear estimation, the nested min–max problem is reformulated using a dummy variable. This variable transforms the original problem into a standard form that can be directly solved using MILP. Thus, (4.13) is rewritten as:

$$\begin{aligned} \min_{\varepsilon,S} \quad & \varepsilon \\ \text{s.t.} \quad & I_n \cdot P_n \cdot \Theta_n \cdot \left(1 - \sum_{j=1}^{M} E_{n,j} \cdot s_j\right) \leq \varepsilon, \quad \forall n = 1, \ldots, N \\ & s.\Phi \leq \mathbf{B} \end{aligned} \tag{4.20}$$

By minimizing ε in problem (4.20), the result guarantees that the maximum risk across all threats remains as low as possible. This formulation leads to an efficient solution using MILP solvers while preserving the risk-aware structure of the original problem.

MaxDegree algorithm

This method considers a weighted bipartite graph $G = (T, C, E)$, where T and C represent disjoint sets of nodes for threats and countermeasures, and E is the weighted adjacency matrix of size $N \times M$ indicating the countermeasure effectiveness matrix.

Each element of $E(n, m)$ represents the weight of the edge between node T_n and node C_m. A weight of zero ($E(n, m) = 0$) indicates the absence of an edge i.e. the countermeasure C_m is not effective for the threat t_n.

MaxDegree initially identifies the nodes in T that can be mitigated by only a single countermeasure. We refer to such nodes as *isolated threats*—that is, t_n connected to exactly one node in C, satisfying the condition

$$|\{m \mid E(n, m) > 0\}| = 1.$$

Considering $S \in \{0, 1\}^{1\times M}$ as the strategy vector which is initially set to all zeros, S_m will be set to 1 for m. Once s_m is activated, all its connected edges must be deactivated. Then, C_m is removed from the set C, and the resulting spanning threats are eliminated from T. Next, the algorithm evaluates the accumulated effectiveness of each remaining countermeasure C_m across all unresolved threats t_n. D_j, the degree of each remaining C_m is calculated as the column-wise sum of the remaining entries in E:

$$D_j = \sum_{i=1}^{N} E(i, m)$$

In the next phase, the algorithm identifies the countermeasure j^* by calculating the maximum D_j and repeats this process until the budget is finished:

$$j^* = \arg\max_j (D_j)$$

MinCost greedy algorithm

The MinCost greedy algorithm inherits the single-node selection approach described in *MaxDegree* algorithm. Initially, the algorithm identifies the isolated threats and activates their corresponding countermeasures C_m by turning on s_m. After this step, the algorithm selects and activates the s_m associated with the C_m that has the minimum cost (provided the budget allows). This process is repeated until no further countermeasures can be added due to budget exhaustion. Figure 4.1 compares the MinCost and MaxDegree methods, illustrating both their common structure and distinct selection strategies.

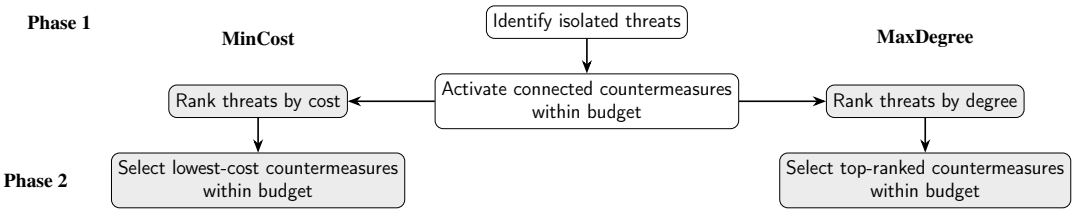

FIGURE 4.1: Comparing two-phase heuristic-based countermeasure selection: MaxDegree and MinCost branches.

MaxThreat algorithm

In contrast to MinCost and MaxDegree, which select countermeasures without explicitly accounting for the attack probability, MaxThreat proposes a complementary

approach that incorporates both the impact of threats and the cost of countermeasures into a unified decision-making process.

MaxThreat begins by evaluating the threat level for each attack $n \in \{1, \ldots, N\}$ based on the following index:

$$n_{\max} = \arg\max_{n} \left(I_n \cdot P_n \cdot \Theta_n\right)$$

To incorporate cost considerations, MaxThreat uses a normalized version of the cost matrix Φ. Each element of Φ is divided by the corresponding budget entry B_j, to reflect the proportion of each countermeasure's cost relative to the available budget. The row-wise summation yields a normalized cost vector:

$$\phi'_i = \sum_{j=1}^{b} \frac{\Phi_{i,j}}{B_j}, \quad \forall i = 1, \ldots, M$$

Next, the effectiveness matrix $E \in \mathbb{R}^{M \times N}$ is adjusted by dividing each element by the corresponding normalized cost, producing the cost-aware effectiveness matrix E_ϕ:

$$(E_\phi)_{i,j} = \frac{E_{i,j}}{\phi'_i}, \quad \forall i = 1, \ldots, M, \quad \forall j = 1, \ldots, N$$

This formulation allows MaxThreat to prioritize countermeasures that offer the highest effectiveness relative to their cost. A higher value in E_ϕ indicates a more cost-efficient and impactful countermeasure.

Finally, MaxThreat selects the countermeasure S_m, where:

$$m = \arg\max_{m} E_\phi[m, n_{\max}]$$

The algorithm then updates the budget and repeats the process for the next most threatening attack $n_{\max}$, continuing until the budget is exhausted.

Genetic algorithm

Similar to other metaheuristics, the GA mimics natural processes to iteratively search for suboptimal solutions to large-scale, intractable mathematical problems. In this chapter, the cost function defined in Eq. (4.13) is provided to the GA in order to find the optimal set S.

4.4 Evaluation

This section presents a series of simulations to evaluate the effectiveness of the proposed methods and analyze their performance in different situations. First, the strategy selection methods are evaluated by applying the core of SASE including 7 main countermeasures [135]. Second, the recommended capabilities of SASE are added, including 13 countermeasures, and finally, the optional capabilities are utilized, including a total of 17 countermeasures. All the simulations have been carried out using a 12th Gen Intel(R) Core(TM) i7-12700 2.10 GHz system.

We adopt $N = 17$ from [155] and consider $M = \{7, 13, 17\}$ as proposed in [135], with $K_T = 5$ and $K_{NT} = 5$ assets located within a specific area served by a designated edge server. We assess the risk and select the optimal set of countermeasures independently for each server. In other words, each edge node evaluates the risk associated with each threat, determines the most suitable set of countermeasures, and applies them locally to the network. Figure 4.2 shows the maximum risk values obtained after applying the proposed countermeasure selection methods. In this setup, K_T, K_{NT}, and N remain constant, while M varies. Moreover, the true probability of attack p_n^R is considered to ensure a fair evaluation of the different optimization methods. Furthermore, a dense matrix E, with uniformly distributed effectiveness values, is used to model countermeasure impacts in this case. As shown in Figure 4.2, the Full search method (i.e., the optimal solution) exhibits decreasing risk as M increases, proving the advantage of using more advanced countermeasure sets. This trend is also observed in MaxThreat, MaxScoring, and GA, indicating their ability to approximate the optimal solution effectively. In contrast, MinCost and MaxDegree show the greatest deviations, as they initially prioritize isolated threat nodes, which may not pose significant risk to the system. The MILP approach also demonstrates some disparity, primarily due to the characteristics of matrix E, which is neither sparse nor composed of small values.

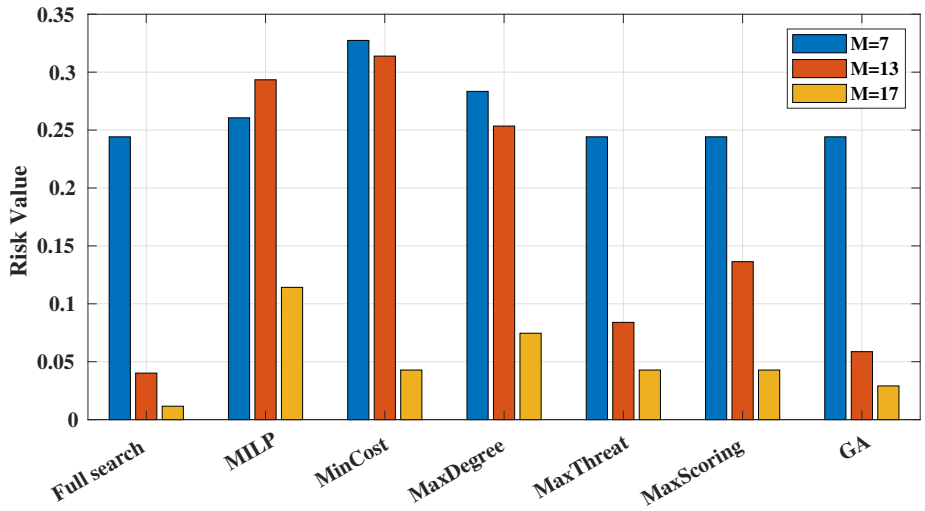

FIGURE 4.2: Comparison of the maximum risk by p_n^R for M=7,13,17 when the strategies selected by each method is applied.

Figure 4.3 compares the maximum risk values obtained from the proposed solutions under the same scenario as in Figure 4.2, considering both the real probability p_n^R and the approximated threat probability p_n. It can be observed that, in this setup, the strategies based on the approximated probability closely align with those based on the real probability across all M values. The only notable disparity in maximum risk values is observed in the case of MILP. This behavior is caused by the branch-and-cut algorithms for solving MILP, where small numerical differences can affect LP relaxation tightness and the integrality of intermediate solutions. As a result, the solver may find a different, suboptimal solution.

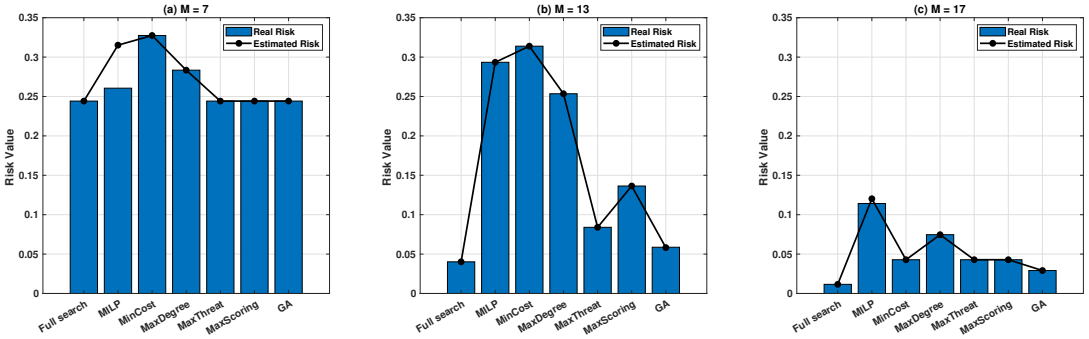

FIGURE 4.3: Comparison of the maximum risk when considering the real and estimated probabilities.

Using the same setup as in Figure 4.2, the implementation times of all algorithms across different values of M are reported in Table 4.2. The GA method is not included in this table due to its very high computational cost. While the algorithm successfully converges to the optimal solution in this scenario, the need to evaluate the fitness function over 400 iterations makes it computationally intensive and less suitable for realtime applications. Among the included methods, Full Search shows a dramatic increase in runtime as M increases. It becomes approximately 130 times slower when M changes from 7 to 17. It is still efficient and provides optimal results for smaller search spaces (e.g., $M = 7$) compared to the other methods. Although it can be a suitable choice for low-dimensional problems, its scalability is clearly limited which underscores the growing importance of faster, suboptimal strategies as the problem size increases. In this context, MaxThreat and MaxScoring exhibit strong performance by consistently providing solutions within significantly shorter timeframes. In contrast, the MILP method—which relies on approximation and the branch-and-cut technique—ranks third in terms of execution time when $M = 17$, and becomes the second least efficient method for $M = 7$ and $M = 13$. These findings highlight the inherent trade-offs between achieving optimality, maintaining computational efficiency, and the speed of the algorithm as the problem scale increases.

TABLE 4.2: Implementation time comparison of the proposed methods.

Method	**Implementation time (s)**		
	$\boldsymbol{M = 7}$	$\boldsymbol{M = 13}$	$\boldsymbol{M = 17}$
Full Search	0.00037	0.00338	0.04871
MILP	0.00945	0.01919	0.01074
MinCost	0.00043	0.00046	0.00049
MaxDegree	0.00032	0.00034	0.00035
MaxThreat	0.00012	0.00013	0.00013
MaxScoring	0.00023	0.00022	0.00026

Apart from the changes in the maximum risk value as M varies, alterations in internal parameters such as the E matrix also lead to different outcomes, especially in MILP. The effectiveness of each countermeasure, reflected in the values of the E matrix, plays a crucial role in selecting the optimal strategy and minimizing the risk.

In Figure 4.4, different models for the E matrix are considered, and the corresponding values are reported for both approximated and real probabilities for $M = 13$. For the approximated probability case, the strategy is first determined using p_n for each method. Then, to ensure a fair comparison, the maximum risk is calculated using Eq. 4.12 by applying p_n^R. According to Figure 4.4 (a) and (c), as the effectiveness of countermeasures increases—reflected by larger values in the E matrix—the overall risk decreases, while the disparity among methods becomes more noticeable. In panel (a), the MinCost method shows the greatest disparity, as it prioritizes the cheapest countermeasures, which may not be reliably effective. Then, the MILP fails in finding the best strategy. This behavior can be attributed to the nature of the linear approximation used in Subsection 4.3.8. Under these conditions, the linear approximation deviates more noticeably from the nonlinear formulation. This is because the linear model, which replaces the product term with a summation, fails to fully capture the compounding effect of multiple strong countermeasures. As the values of E increase, the product term in the original nonlinear function can drop significantly below the value estimated by the linear approximation. As a result, the linear model overestimates the risk that leads to a gap between MILP and Full search. This divergence highlights the limitations of the linear approximation when countermeasures are highly effective. In addition, the risk values in panel (d) (corresponding to the sparse and small E case) completely converge to a high maximum risk value for both p_n and p_n^R across all methods. This supports the observation that when the effectiveness of countermeasures is very low and E contains many zeros, the entire system becomes ineffective.

Figure 4.5 illustrates the maximum risk evaluated using p_n across the proposed methods. In this simulation, the most advanced version of the SASE framework is employed, configured with $M = 17$. The matrix E is designed as a sparse matrix representing effective countermeasures, closely resembling real-world scenarios. The window length is set to $T_f = 334$ timeslots, where initially the system operates solely with $p_n^E = 0.3$. Once sufficient information is collected after T_f, the attack probability p_n^M is incorporated alongside p_n^E to from p_n. Since $p_n = p_n^E$ and remains constant for $t < T_f$, the estimated maximum risk does not change during that period. After T_f, the estimated maximum risk values change in accordance to p_n and the corresponding selected strategy. Owing to the specific configuration of E, the MILP solution aligns closely with the optimal result, while the MinCost approach exhibits the largest deviation. The MaxScoring method ranks second in accuracy, following the MILP, in approximating the optimal solution. In addition, GA demonstrates strong alignment with the optimal solution for the majority of the simulation period; however, it occasionally becomes trapped in local minima, leading to fluctuations between suboptimal and near-optimal performance. Figure 4.6 clearly illustrates the differences between the optimal solution and the suboptimal ones in terms of error percentage. This allows the enterprise owner to assess their risk appetite and select the most suitable risk treatment method based on the error levels of the proposed approaches.

Moreover, Figure 4.7 illustrates the Hamming distance among different solutions across timeslots. Although the GA converges well to the optimal solution, the Hamming distance remains around 5, with occasional fluctuations up to 7. This indicates that when there are 2^{17} possible solutions for the S vector, even strategies that are relatively distant in solution space can yield suitable sub-optimal results for risk minimization. In contrast, the conditions were useful for MILP in enabling accurate risk

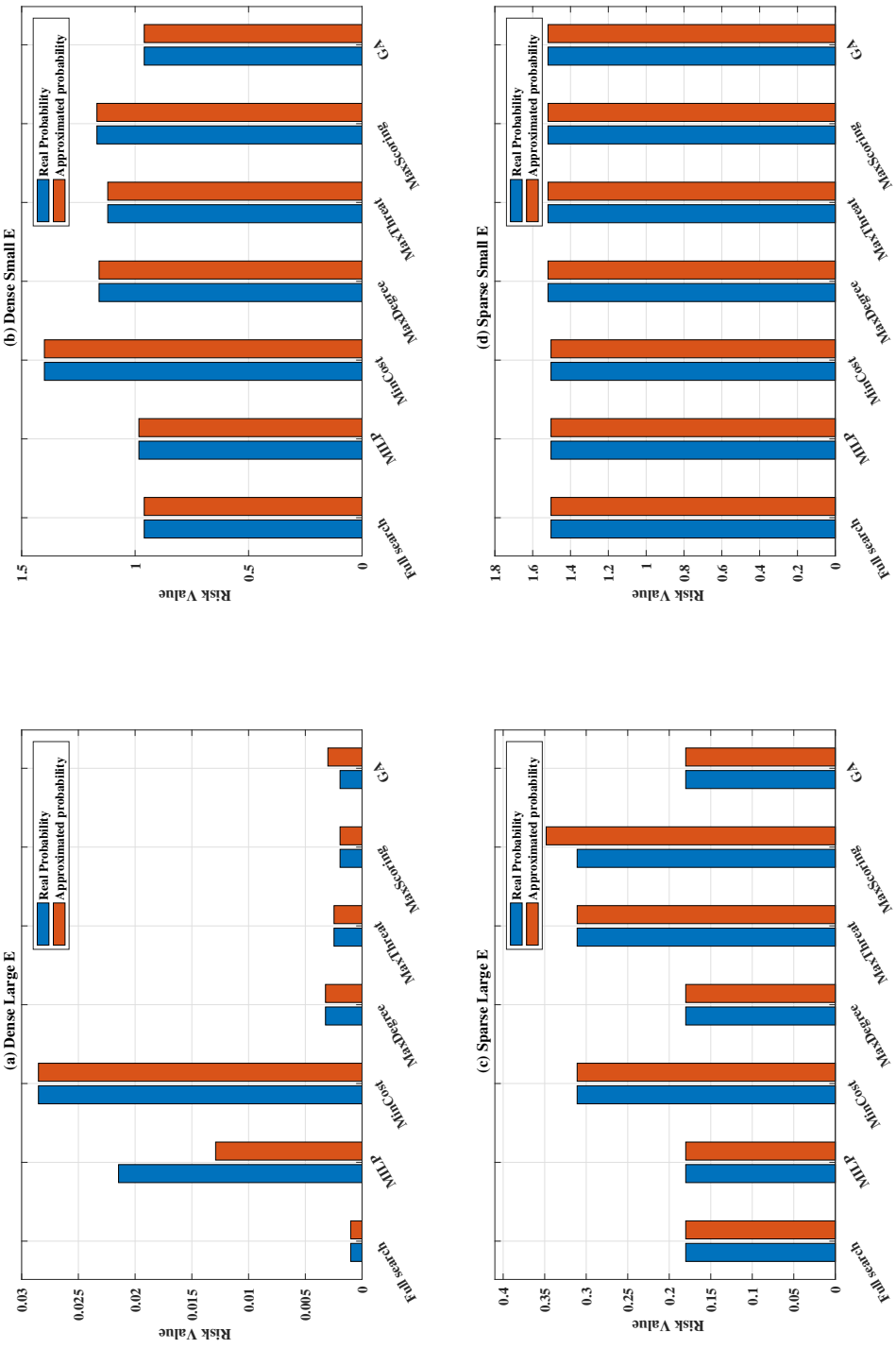

FIGURE 4.4: Maximum risk value for M=13 evaluated when various E matrices are considered.

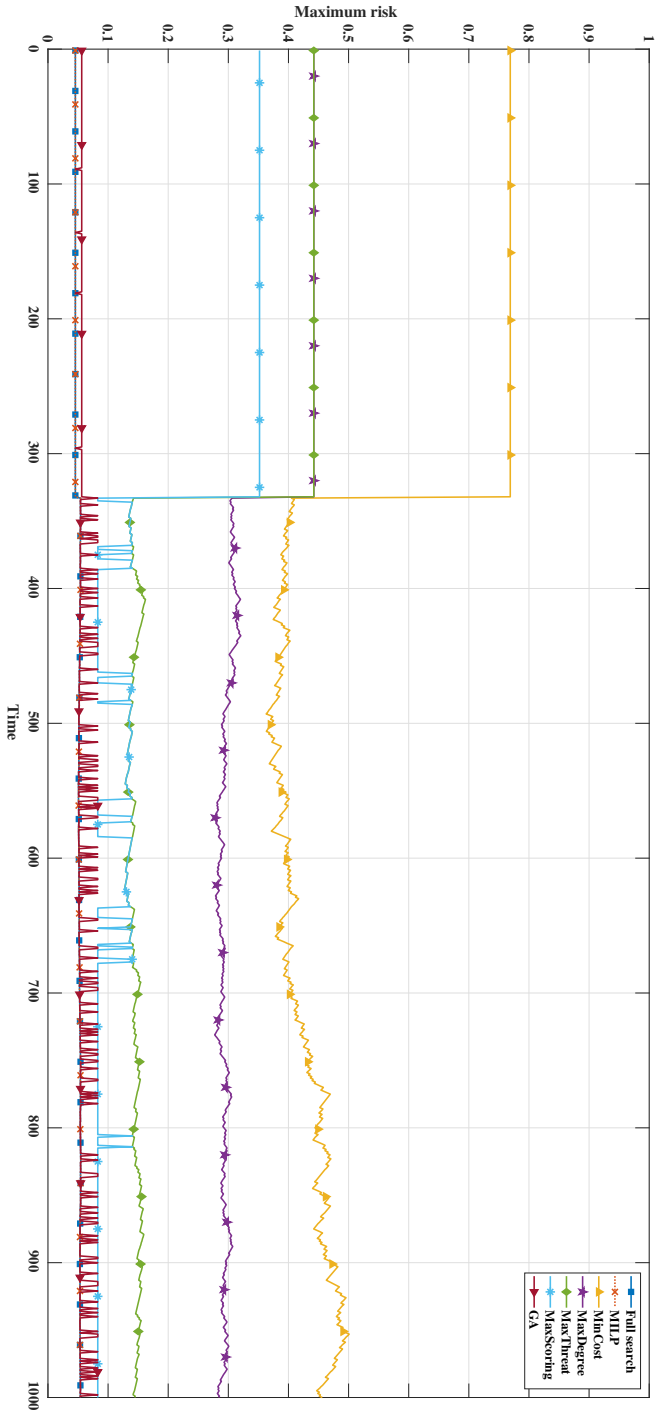

FIGURE 4.5: Maximum risk value for M=17 in 1000 timeslots using p_n.

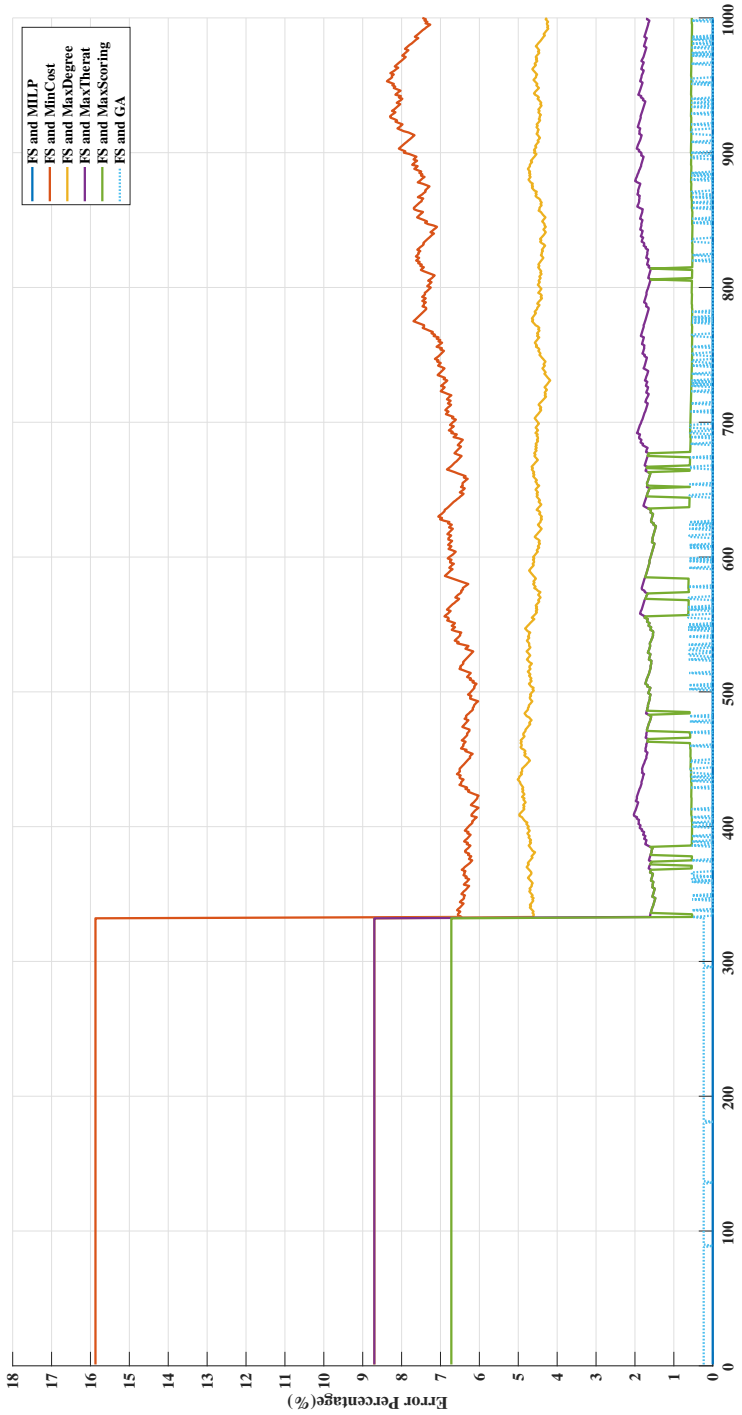

FIGURE 4.6: Error percentage of each method compared to the Full search result in 1000 timeslots using p_n.

estimation and determining the S vector with minimal error and Hamming distance. The steady trends of MinCost and MaxDegree stem from the way they select countermeasures, which only consider the impact and vulnerability assessment steps. Their selection is independent of attack probabilities, meaning that the S vector remains unchanged regardless of the likelihood of attack.

4.5 Conclusions

In this chapter, a novel risk-based adaptive security framework tailored for 6G edge computing environments is presented which integrates the SASE architecture and the IRAM2 risk assessment methodology. The framework models cyber risk through a detailed analysis of threat impact, attack probabilities, countermeasure effectiveness, and various treatment strategies and yields dynamic defense against potential threats. The risk treatment process was formulated as a constrained optimization problem in which various strategies, including exact, approximate, and heuristic algorithms were developed and evaluated. Extensive simulations demonstrated that while full search and GA achieve optimal risk reduction, they suffer from high computational costs. In contrast, the proposed heuristics, particularly MaxScoring and MaxThreat, offered a favorable trade-off between accuracy and computational efficiency, making them suitable for real-time computations in the edge. Moreover, the effectiveness of linear approximation via MILP was found to be sensitive to the structure and sparsity of E matrix. Overall, this work highlights the importance of resource-efficient and reliable frameworks in next generation mobile networks. Future work will investigate distributed coordination among edge nodes, consider evolving threat models and users with special budgets, and explore reinforcement learning-based adaptive security strategies to further enhance the robustness of the proposed framework. Moreover, this chapter introduced SASE as a security solution for NTN and UAV networks. While SASE provides a useful baseline, current SASE offerings are not sufficient to fully address the requirements of NTNs and UAV networks. Future work should focus on extending SASE architectures with NTN-aware capabilities to achieve more complete protection.

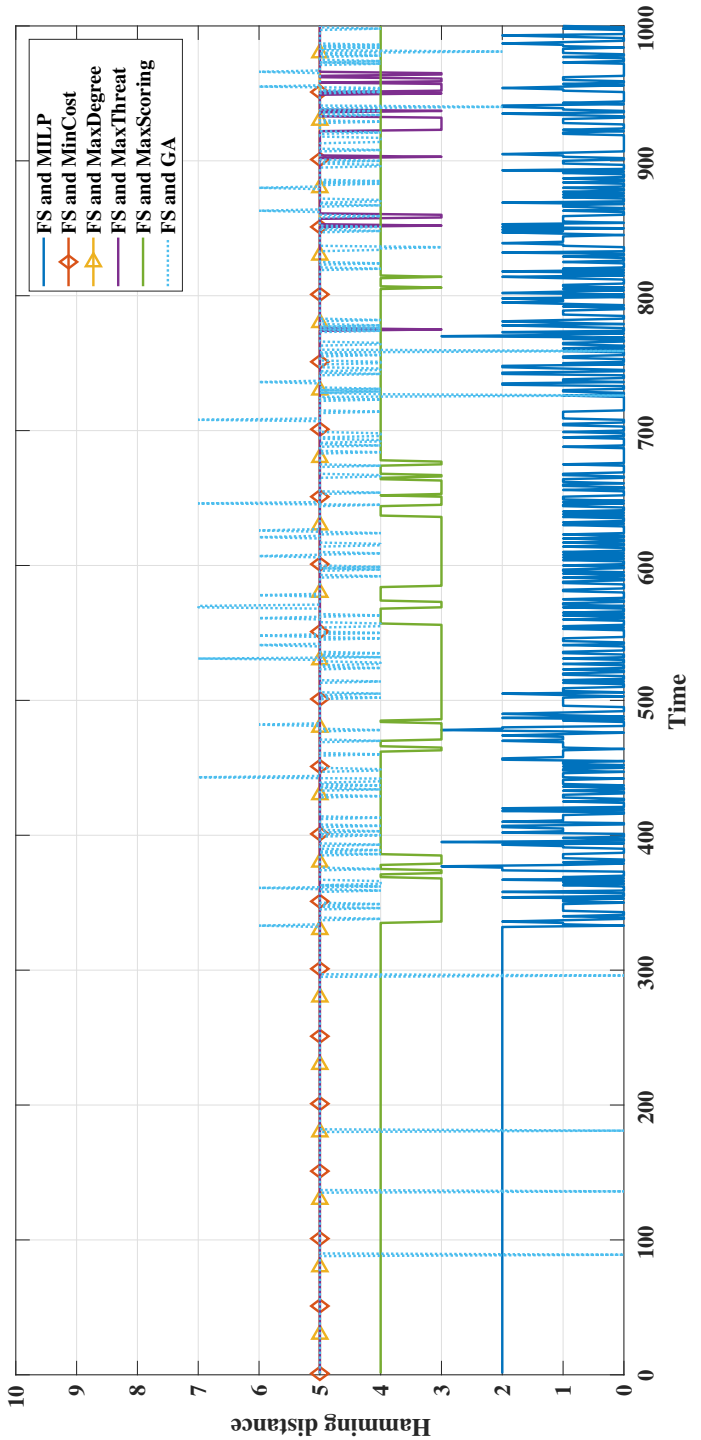

FIGURE 4.7: Hamming distance between the selected strategy by the Full search and other methods in 1000 timeslots using p_n.

Conclusions and Future Works

Integration of TN and NTNs is a key step toward the realization of 6G systems, allowing seamless connectivity across the space, air, and ground segments. However, this integration introduces significant challenges that cannot be fully addressed by 5G technologies, including limited communication resources, noisy and adverse channel conditions, and an expanded attack surface due to the broadcast nature of wireless transmission and the inclusion of AI. Addressing these challenges requires communication and security mechanisms that are not only efficient, but also robust, adaptive to heterogeneous and resource-constrained environments, and capable of supporting low-latency requirements. This thesis has addressed the mentioned issues through the design of complementary frameworks that target different aspects of the problem, ranging from transmission efficiency to communication security and system-level risk management.

First, this thesis explored how communication efficiency can be improved in noisy wireless channels by leveraging task-oriented transmission principles. A lightweight JSCC-based framework was developed for video transmission, integrating semantic awareness through attention mechanisms and SNR-adaptive modulation. This approach prioritizes task-relevant information which results in improved video key frame reconstruction on the receiver's side, adapt to channel conditions, compress the video according to the specific task, particularly under low-SNR conditions and strict bandwidth constraints.

Second, this thesis investigated the security vulnerabilities existing in SemCom systems applicable to TN-NTN scenarios, particularly those leveraging AI to encode data for wireless transmission. The analysis examined the main threats targeting SemCom systems, with a focus on eavesdropping and privacy preservation in wiretap channels. Existing methods for countering eavesdropping were reviewed, and the limitations of conventional protection strategies were highlighted when applied to SemCom transmission schemes. In addition, a comprehensive comparison of state-of-the-art approaches against MIEA, along with a newly proposed hybrid method was presented. To further address the MIEA issue, a lightweight and reversible protection mechanism (KAP-DJSCC) was introduced. The KAP-DJSCC scheme secures latent representations through key-based transformations, effectively reducing the ability of an eavesdropper to reconstruct meaningful information while preserving performance for legitimate users. These results demonstrate that privacy can be enhanced without introducing significant computational overhead or degrading communication quality.

Finally, the thesis extended beyond communication-level solutions to address broader cybersecurity challenges in 6G systems. A risk-based adaptive security framework was introduced to support dynamic decision-making in the presence of multiple threats, resource limitations, and heterogeneous system requirements. Unlike static approaches, the proposed framework statistically evaluates risk in real-time and selects appropriate countermeasures based on system conditions and constraints.

This allows for a more flexible and scalable security strategy, applicable to a wide range of TN–NTN scenarios.

Overall, these contributions show that the challenges of TN–NTN integration cannot be addressed by a single solution. Achieving efficient and secure communication in 6G systems requires a combination of advanced transmission techniques, lightweight and reliable protection mechanisms, and adaptive system-level strategies that are tailored to the requirements of specific applications. Looking ahead, several directions can be explored to further enhance and extend the proposed frameworks. In particular, future work will focus on addressing the limitations identified in this thesis and moving toward more realistic and application-oriented scenarios.

For instance, in SemCom-based NTN systems, more practical channel models and latency constraints should be considered beyond the AWGN assumption. The proposed architectures can be extended to account for the unique characteristics of NTN environments, such as long-distance satellite links and the mobility of UAVs. Incorporating these factors would enable the development of more robust and tailored semantic image and video transmission schemes. Moreover, the dynamic nature of wireless channels in TN–NTN systems requires communication models that can adapt to sudden variations in channel conditions. In this context, techniques such as transfer learning and generative modeling can be leveraged to improve the adaptability of the proposed SemCom frameworks, particularly JSCC-based approaches. These methods can provide proactive training and enhance the resilience of the system against rapidly changing channel environments.

In free-space environments, optical communication can support high data rates and large bandwidth while maintaining low size, weight, and power requirements. Integrating free space communication with SemCom provides a promising direction for efficient and task-oriented transmission in TN–NTN systems. However, challenges such as atmospheric effects, alignment issues, and security risks remain critical [156]. Future work can focus on extending the proposed security mechanisms to free space semantic communication, particularly through data-driven protection and semantic-level covertness. In addition, the privacy-utility trade-off approaches can be explored to improve security and robustness under dynamic channel conditions, while maintaining high-quality image and video transmission.

In addition, in air-to-ground scenarios, the wireless channel exhibits unique characteristics with random variations in CSI. These variations can be exploited to generate shared cryptographic keys between legitimate users, supporting PLS against eavesdroppers. In particular, channel effects such as fading, noise, and Doppler shifts can be leveraged to enhance communication security. Beyond the use of deep encoders and decoders for data transmission in NTN, future research can focus on designing frameworks that explicitly incorporate physical layer properties of NTNs to achieve more secure and robust transmission.

Bibliography

[1] H. Peng, Z. Zhang, Y. Liu, Z. Su, T. H. Luan, and N. Cheng, "Semantic communication in non-terrestrial networks: A future-ready paradigm," *IEEE network*, vol. 38, no. 4, pp. 119–127, 2024.

[2] M. M. Azari et al., "Evolution of non-terrestrial networks from 5g to 6g: A survey," *IEEE communications surveys & tutorials*, vol. 24, no. 4, pp. 2633–2672, 2022.

[3] Y. Qu et al., "Empowering edge intelligence by air-ground integrated federated learning," *IEEE Network*, vol. 35, no. 5, pp. 34–41, 2021.

[4] A. U. Haq, S. S. Sefati, S. J. Nawaz, A. Mihovska, and M. J. Beliatis, "Need of uavs and physical layer security in next-generation non-terrestrial wireless networks: Potential challenges and open issues," *IEEE Open Journal of Vehicular Technology*, vol. 6, pp. 554–595, 2025.

[5] G. Geraci, D. López-Pérez, M. Benzaghta, and S. Chatzinotas, "Integrating terrestrial and non-terrestrial networks: 3d opportunities and challenges," *IEEE Communications Magazine*, vol. 61, no. 4, pp. 42–48, 2022.

[6] D.-H. Jung, J.-G. Ryu, and J. Choi, "When satellites work as eavesdroppers," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2784–2799, 2022.

[7] M. A. Jamshed et al., "Non-terrestrial networks for 6g: Integrated, intelligent and ubiquitous connectivity," *IEEE Communications Standards Magazine*, 2025.

[8] S. Guo et al., "A survey on semantic communication networks: Architecture, security, and privacy," *IEEE Communications Surveys & Tutorials*, 2024.

[9] E. Bourtsoulatze, D. B. Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567–579, 2019.

[10] M. N. Kenari, G. Sciddurlo, L. A. Grieco, and N. Cordeschi, "Privacy preserving in wiretap channels beyond shannon's paradigm: Current trends and future challenges,"

[11] S. Shen, C. Yu, K. Zhang, J. Ni, and S. Ci, "Adaptive and dynamic security in ai-empowered 6g: From an energy efficiency perspective," *IEEE Communications Standards Magazine*, vol. 5, no. 3, pp. 80–88, 2021.

[12] M. N. Kenari, A. Fascista, I. Cianci, G. Sciddurlo, L. A. Grieco, and N. Cordeschi, "Risk-based adaptive security framework for 6g edge,"

[13] Z. Zhang, M. Debbah, Y. C. Eldar, D. T. Hoang, W. Tong, and K.-K. Wong, "Guest editorial: Sustainable big ai model for wireless networks," *IEEE Wireless Communications*, vol. 31, no. 3, pp. 18–19, 2024.

[14] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[15] C. E. Shannon and W. Weaver, "The mathematical theory of communication. university of illinois," *Urbana*, vol. 117, no. 10, 1949.

[16] Z. Qin, X. Tao, J. Lu, W. Tong, and G. Y. Li, "Semantic communications: Principles and challenges," *arXiv preprint arXiv:2201.01389*, 2021.

[17] D. Gündüz et al., "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 5–41, 2022.

[18] E. C. Strinati et al., "Goal-oriented and semantic communication in 6g ai-native networks: The 6g-goals approach," in *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 2024, pp. 1–6. DOI: 10.1109/EuCNC/6GSummit60053.2024.10597087.

[19] A. Mostaani, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Task-effective compression of observations for the centralized control of a multi-agent system over bit-budgeted channels," *IEEE Internet of Things Journal*, 2023.

[20] W. Yang et al., "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 213–250, 2022.

[21] Y. E. Sagduyu, T. Erpek, A. Yener, and S. Ulukus, "Will 6g be semantic communications? opportunities and challenges from task oriented and secure communications to integrated sensing," *IEEE Network*, 2024.

[22] E. C. Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107 930, 2021.

[23] E. C. Strinati et al., "Goal-oriented and semantic communication in 6g ai-native networks: The 6g-goals approach," in *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, IEEE, 2024, pp. 1–6.

[24] J. Wu et al., "Semantic segmentation-based semantic communication system for image transmission," *Digital Communications and Networks*, vol. 10, no. 3, pp. 519–527, 2024, ISSN: 2352-8648. DOI: https://doi.org/10.1016/j.dcan.2023.02.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235286482300041X.

[25] H. He, S. Lyu, and B. Feng, "Secure transmission over multiple access wiretap channel by cross-time interference injection," *IEEE Transactions on Communications*, vol. 71, no. 1, pp. 370–380, 2022.

[26] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, "Hardware-impaired phy secret key generation with man-in-the-middle adversaries," *IEEE Wireless Communications Letters*, vol. 11, no. 4, pp. 856–860, 2022.

[27] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, "Deep learning for hardware-impaired wireless secret key generation with man-in-the-middle attacks," in *2021 IEEE Global Communications Conference (GLOBECOM)*, IEEE, 2021, pp. 1–6.

[28] X. Liu et al., "Semprotector: A unified framework for semantic protection in deep learning-based semantic communication systems," *IEEE Communications Magazine*, vol. 61, no. 11, pp. 56–62, 2023.

[29] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *2017 IEEE symposium on security and privacy (SP)*, IEEE, 2017, pp. 3–18.

[30] B. Yang et al., "Reconfigurable intelligent computational surfaces: When wave propagation control meets computing," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 120–128, 2023.

[31] Y. Wang, W. Yang, P. Guan, Y. Zhao, and Z. Xiong, "Star-ris-assisted privacy protection in semantic communication system," *IEEE Transactions on Vehicular Technology*, 2024.

[32] J. Chen, J. Wang, C. Jiang, Y. Ren, and L. Hanzo, "Trustworthy semantic communications for the metaverse relying on federated learning," *IEEE Wireless Communications*, vol. 30, no. 4, pp. 18–25, 2023.

[33] Y. Yang, M. Shikh-Bahaei, Z. Yang, C. Huang, W. Xu, and Z. Zhang, "Joint semantic communication and target sensing for 6g communication system," *arXiv preprint arXiv:2401.17108*, 2024.

[34] X. Mu and Y. Liu, "Semantic communication-assisted physical layer security over fading wiretap channels," in *ICC 2024-IEEE International Conference on Communications*, IEEE, 2024, pp. 2101–2106.

[35] T.-Y. Tung and D. Gündüz, "Deep joint source-channel and encryption coding: Secure semantic communications," in *ICC 2023-IEEE International Conference on Communications*, IEEE, 2023, pp. 5620–5625.

[36] R. Zhao, Q. Qin, N. Xu, G. Nan, Q. Cui, and X. Tao, "Semkey: Boosting secret key generation for ris-assisted semantic communication systems," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, IEEE, 2022, pp. 1–5.

[37] Y. Chen, Q. Yang, Z. Shi, and J. Chen, "The model inversion eavesdropping attack in semantic communication systems," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*, IEEE, 2023, pp. 5171–5177.

[38] F. Chen, L. Xiang, H. V. Cheng, and K. Shen, "Deep learning enabled semantic-secure communication with shuffling," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*, IEEE, 2023, pp. 6338–6343.

[39] T. Guo, J. Han, H. Wu, Y. Wang, B. Bai, and W. Han, "Protecting semantic information using an efficient secret key," in *2022 IEEE International Symposium on Information Theory (ISIT)*, IEEE, 2022, pp. 2660–2665.

[40] X. Luo, Z. Chen, M. Tao, and F. Yang, "Encrypted semantic communication using adversarial training for privacy preserving," *IEEE Communications Letters*, 2023.

[41] T. Marchioro, N. Laurenti, and D. Gündüz, "Adversarial networks for secure wireless communications," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2020, pp. 8748–8752.

[42] J. Xu, B. Ai, W. Chen, N. Wang, and M. Rodrigues, "Deep joint source-channel coding for image transmission with visual protection," *IEEE Transactions on Cognitive Communications and Networking*, 2023.

[43] L. Sun, C. Guo, M. Chen, and Y. Yang, "Privacy-aware joint source-channel coding for image transmission based on disentangled information bottleneck," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2024, pp. 9016–9020.

[44] 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.

[45] S. A. A. Kalkhoran, M. Letafati, E. Erdemir, B. H. Khalaj, H. Behroozi, and D. Gündüz, "Secure deep-jscc against multiple eavesdroppers," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*, IEEE, 2023, pp. 3433–3438.

[46] M. Zhang, Y. Li, Z. Zhang, G. Zhu, and C. Zhong, "Wireless image transmission with semantic and security awareness," *IEEE Wireless Communications Letters*, 2023.

[47] S. Tang, C. Liu, Q. Yang, S. He, and D. Niyato, "Secure semantic communication for image transmission in the presence of eavesdroppers," *arXiv preprint arXiv:2404.12170*, 2024.

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

[49] Y. Zhang, X. Yuan, W. Li, and S. Chen, "Automatic power line inspection using uav images," *Remote Sensing*, vol. 9, no. 8, p. 824, 2017.

[50] G. Shi, Y. Xiao, Y. Li, and X. Xie, "From semantic communication to semantic-aware networking: Model, architecture, and open problems," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 44–50, 2021. DOI: 10.1109/MCOM.001.2001239.

[51] Q. Lan et al., "What is semantic communication? a view on conveying meaning in the era of machine intelligence," *Journal of Communications and Information Networks*, vol. 6, no. 4, pp. 336–371, 2021.

[52] T. M. Getu, G. Kaddoum, and M. Bennis, "Semantic communication: A survey on research landscape, challenges, and future directions," *Proceedings of the IEEE*, 2025.

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

[54] K. Lu et al., "Rethinking modern communication from semantic coding to semantic communication," *IEEE Wireless Communications*, vol. 30, no. 1, pp. 158–164, 2022.

[55] M. Shen et al., "Secure semantic communications: Challenges, approaches, and opportunities," *IEEE Network*, 2023.

[56] H. Du et al., "Rethinking wireless communication security in semantic internet of things," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 36–43, 2023.

[57] R. Meng et al., "A survey of secure semantic communications," *Journal of Network and Computer Applications*, p. 104 181, 2025.

[58] B. Wang et al., "Image steganography for securing intellicise wireless networks:" invisible encryption" against eavesdroppers," *arXiv preprint arXiv:2505.04467*, 2025.

[59] S. Cheng, X. Zhang, Y. Sun, Q. Cui, and X. Tao, "Knowledge discrepancy oriented privacy preserving for semantic communication," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 8, pp. 11 637–11 646, 2024. DOI: 10.1109/TVT.2024.3381222.

[60] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.

[61] H. Xiao et al., "Star-ris-assisted joint physical layer security and covert communications," in *2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall)*, 2023, pp. 1–6. DOI: 10.1109/VTC2023-Fall60731.2023.10333375.

[62] X. Chen et al., "Covert communications: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1173–1198, 2023.

[63] T. Chapman, E. Larsson, P. von Wrycza, E. Dahlman, S. Parkvall, and J. Sköld, "Chapter 3 - cdma transmission principles," in *HSPA Evolution*, T. Chapman, E. Larsson, P. von Wrycza, E. Dahlman, S. Parkvall, and J. Sköld, Eds., Oxford: Academic Press, 2015, pp. 35–48, ISBN: 978-0-08-099969-2. DOI: https://doi.org/10.1016/B978-0-08-099969-2.00003-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978008099969200003X.

[64] Y. Yao, Z. Xu, and W. Li, "Visual security assessment for video encryption," in *2008 Third International Conference on Communications and Networking in China*, 2008, pp. 1317–1322. DOI: 10.1109/CHINACOM.2008.4685269.

[65] O. F. Mohammad, M. S. M. Rahim, S. R. M. Zeebaree, and F. Ahmed, "A survey and analysis of the image encryption methods," *International Journal of Applied Engineering Research*, vol. 12, no. 23, pp. 13 265–13 280, 2017.

[66] T. Xiao, Y.-H. Tsai, K. Sohn, M. Chandraker, and M.-H. Yang, "Adversarial learning of privacy-preserving and task-oriented representations," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 2020, pp. 12 434–12 441.

[67] E. Erdemir, P. L. Dragotti, and D. Gündüz, "Privacy-aware communication over a wiretap channel with generative networks," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2022, pp. 2989–2993.

[68] Y. Sun, H. Wang, L. Wang, B. Hu, and W. Wu, "Multi-irss enhanced uav secure semantic communication system," *Physical Communication*, p. 102 427, 2024.

[69] W. Chen, S. Shao, Q. Yang, Z. Zhang, and P. Zhang, "A nearly information theoretically secure approach for semantic communications over wiretap channel," *arXiv preprint arXiv:2401.13980*, 2024.

[70] L. Sun, Y. Yang, M. Chen, and C. Guo, "Disentangled information bottleneck guided privacy-protective joint source and channel coding for image transmission," *IEEE Transactions on Communications*, 2024.

[71] G. Chen et al., "Lightweight and robust wireless semantic communications," *IEEE Communications Letters*, 2024.

[72] B. He, F. Wang, and T. Q. Quek, "Secure semantic communication via paired adversarial residual networks," *IEEE Wireless Communications Letters*, 2024.

[73] M. Costa and Y. E. Sagduyu, "Timely and covert communications under deep learning-based eavesdropping and jamming effects," *Journal of Communications and Networks*, vol. 25, no. 5, pp. 621–630, 2023.

[74] 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.

[75] Y. Liu, J. Wen, Z. Zhang, K. Zhu, and J. Kang, "Learning-based power control for secure covert semantic communication," *arXiv preprint arXiv:2407.07475*, 2024.

[76] 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 *2024 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2024, pp. 1–6.

[77] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2663–2675, 2021.

[78] Q. Qin et al., "Securing semantic communications with physical-layer semantic encryption and obfuscation," in *ICC 2023-IEEE International Conference on Communications*, IEEE, 2023, pp. 5608–5613.

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

[80] R. Kaewpuang et al., "Cooperative resource management in quantum key distribution (qkd) networks for semantic communication," *IEEE Internet of Things Journal*, 2023.

[81] G. Li, Y. Zhao, and Y. Li, "Catfl: Certificateless authentication-based trustworthy federated learning for 6g semantic communications," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, IEEE, 2023, pp. 1–6.

[82] W. Wei, L. Liu, Y. Wu, G. Su, and A. Iyengar, "Gradient-leakage resilient federated learning," in *2021 IEEE 41st International Conference on Distributed Computing Systems (ICDCS)*, IEEE, 2021, pp. 797–807.

[83] L. Tong, F. Dai, Y. Zhang, and J. Li, "Visual security evaluation for video encryption," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 835–838.

[84] X. Liu et al., "Knowledge-assisted privacy preserving in semantic communication," *arXiv preprint arXiv:2410.18418*, 2024.

[85] A. Bounhar, M. Sarkiss, and M. Wigger, "Unveiling covert semantics: Joint source-channel coding under a covertness constraint," *arXiv preprint arXiv:2406.13466*, 2024.

[86] A. Y. Huang, Y. Chen, D. Huang, and M. Zhao, "Semantic privacy-preserving for video surveillance services on the edge," in *2023 IEEE/ACM Symposium on Edge Computing (SEC)*, IEEE, 2023, pp. 300–305.

[87] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, "On learning-assisted content-based secure image transmission for delay-aware systems with randomly-distributed eavesdroppers," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1125–1139, 2021.

[88] M. Letafati, H. Behroozi, B. H. Khalaj, and E. A. Jorswieck, "Content-based medical image transmission against randomly-distributed passive eavesdroppers," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, IEEE, 2021, pp. 1–7.

[89] V. N. I. Cisco, "Cisco visual networking index: Forecast and methodology, 2016–2021," *CISCO White paper*, vol. 2022, 2017.

[90] T.-Y. Tung and D. Gündüz, "Deepwive: Deep-learning-aided wireless video transmission," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2570–2583, 2022.

[91] D. Huang, F. Gao, X. Tao, Q. Du, and J. Lu, "Toward semantic communications: Deep learning-based image semantic coding," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 55–71, 2022.

[92] F. Z. Safaeipour and M. Hashemi, "Semantic-aware and goal-oriented communications for object detection in wireless end-to-end image transmission," *arXiv preprint arXiv:2402.01064*, 2024. DOI: 10.48550/arXiv.2402.01064. [Online]. Available: https://arxiv.org/abs/2402.01064.

[93] S. Wang et al., "Wireless deep video semantic transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 214–229, 2023. DOI: 10.1109/JSAC.2022.3221977.

[94] Q. Du, Y. Duan, Q. Yang, X. Tao, and M. Debbah, "Object-attribute-relation representation based video semantic communication," *IEEE Journal on Selected Areas in Communications*, 2025.

[95] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Wireless semantic communications for video conferencing," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 230–244, 2023. DOI: 10.1109/JSAC.2022.3221968.

[96] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703.

[97] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.

[98] Y. Yuan, X. Chen, and J. Wang, "Object-contextual representations for semantic segmentation," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*, Springer, 2020, pp. 173–190.

[99] T.-C. Wang et al., "Video-to-video synthesis," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[100] M. Yang and H.-S. Kim, "Deep joint source-channel coding for wireless image transmission with adaptive rate control," *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5193–5197, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:238583132.

[101] A. Vaswani et al., "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[102] J. Fu et al., "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 3146–3154.

[103] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[104] M. Cordts et al., "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[105] K. D. B. J. Adam et al., "A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, vol. 1412, no. 6, 2014.

[106] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 586–595.

[107] Y. Yuan, X. Chen, X. Chen, and J. Wang, *Segmentation transformer: Object-contextual representations for semantic segmentation*, 2021. arXiv: 1909.11065 [cs.CV]. [Online]. Available: https://arxiv.org/abs/1909.11065.

[108] F. Bellard. "Better portable graphics." Available at: https://bellard.org/bpg/. [Online]. Available: https://bellard.org/bpg/.

[109] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

[110] M. Yang, C. Bian, and H.-S. Kim, "Deep joint source channel coding for wireless image transmission with ofdm," in *ICC 2021-IEEE International Conference on Communications*, IEEE, 2021, pp. 1–6.

[111] Z. Li, J. Zhou, G. Nan, Z. Li, Q. Cui, and X. Tao, "Sembat: Physical layer black-box adversarial attacks for deep learning-based semantic communication systems," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, IEEE, 2022, pp. 1–5.

[112] X. Yuan, P. He, Q. Zhu, and X. Li, "Adversarial examples: Attacks and defenses for deep learning," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2805–2824, 2019.

[113] Z. Che et al., "Adversarial attack against deep saliency models powered by non-redundant priors," *IEEE Transactions on Image Processing*, vol. 30, pp. 1973–1988, 2021.

[114] A. Krizhevsky, G. Hinton, et al., "Learning multiple layers of features from tiny images," 2009.

[115] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, Ieee, vol. 2, 2003, pp. 1398–1402.

[116] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.

[117] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[118] S. Tang, Y. Chen, Q. Yang, R. Zhang, D. Niyato, and Z. Shi, "Towards secure semantic communications in the presence of intelligent eavesdroppers," *arXiv preprint arXiv:2503.23103*, 2025.

[119] M. Letafati, S. A. A. Kalkhoran, E. Erdemir, B. H. Khalaj, H. Behroozi, and D. Gündüz, "Deep joint source channel coding for privacy-aware end-to-end image transmission," *IEEE Transactions on Machine Learning in Communications and Networking*, 2025.

[120] M. E. O'neill, "Pcg: A family of simple fast space-efficient statistically good algorithms for random number generation," *ACM Transactions on Mathematical Software*, vol. 204, pp. 1–46, 2014.

[121] S. V. Dibbo, "Sok: Model inversion attack landscape: Taxonomy, challenges, and future roadmap," in *2023 IEEE 36th Computer Security Foundations Symposium (CSF)*, IEEE, 2023, pp. 439–456.

[122] A. Zhang, Y. Wang, and S. Guo, "On the utility-informativeness-security trade-off in discrete task-oriented semantic communication," *IEEE Communications Letters*, 2024.

[123] T. M. Getu, G. Kaddoum, and M. Bennis, "Making sense of meaning: A survey on metrics for semantic and goal-oriented communication," *IEEE Access*, vol. 11, pp. 45 456–45 492, 2023.

[124] Y. Lin et al., "Blockchain-aided secure semantic communication for ai-generated content in metaverse," *IEEE Open Journal of the Computer Society*, vol. 4, pp. 72–83, 2023.

[125] B. Duo, J. Luo, Y. Li, H. Hu, and Z. Wang, "Joint trajectory and power optimization for securing uav communications against active eavesdropping," *China Communications*, vol. 18, no. 1, pp. 88–99, 2021.

[126] Y. Wang et al., "Large model based agents: State-of-the-art, cooperation paradigms, security and privacy, and future trends," *IEEE Communications Surveys & Tutorials*, 2025.

[127] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang, and D. Zhang, "A survey on green 6g network: Architecture and technologies," *IEEE Access*, vol. 7, pp. 175 758–175 768, 2019. DOI: 10.1109/ACCESS.2019.2957648.

[128] Q. Bi, "Ten trends in the cellular industry and an outlook on 6g," *IEEE Communications Magazine*, vol. 57, no. 12, pp. 31–36, 2019. DOI: 10.1109/MCOM.001.1900315.

[129] B. Mao, J. Liu, Y. Wu, and N. Kato, "Security and privacy on 6g network edge: A survey," *IEEE Communications Surveys & Tutorials*, 2023.

[130] P. Porambage, G. Gür, D. P. M. Osorio, M. Liyanage, A. Gurtov, and M. Ylianttila, "The roadmap to 6g security and privacy," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1094–1122, 2021.

[131] S. Shen, C. Yu, K. Zhang, J. Ni, and S. Ci, "Adaptive and dynamic security in ai-empowered 6g: From an energy efficiency perspective," *IEEE Communications Standards Magazine*, vol. 5, no. 3, pp. 80–88, 2021. DOI: 10.1109/MCOMSTD.101.2000090.

[132] *Nist glossary*, https://csrc.nist.gov/glossary/term/cybersecurity_risk.

[133] C. Tikkinen-Piri, A. Rohunen, and J. Markkula, "Eu general data protection regulation: Changes and implications for personal data collecting companies," *Computer Law & Security Review*, vol. 34, no. 1, pp. 134–153, 2018.

[134] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Generation Computer Systems*, vol. 97, pp. 219–235, 2019.

[135] N. MacDonald, L. Orans, and J. Skorupa, "The future of network security is in the cloud," *Gartner, htt ps://www. gartner. com/doc/reprints*, 2019.

[136] S. van der Walt and H. Venter, "Research gaps and opportunities for secure access service edge," in *International Conference on Cyber Warfare and Security*, vol. 17, 2022, pp. 609–619.

[137] Palo Alto Networks, *What Is SASE (Secure Access Service Edge)?* https://www.paloaltonetworks.com/cyberpedia/what-is-sase, [Accessed: 2025-06-18], 2025.

[138] S. Hong, L. Xu, J. Huang, H. Li, H. Hu, and G. Gu, "Sysflow: Toward a programmable zero trust framework for system security," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2794–2809, 2023. DOI: 10.1109/TIFS.2023.3264152.

[139] P. He et al., "Non-terrestrial network technologies: Applications and future prospects," *IEEE Internet of Things Journal*, 2024.

[140] R. Sato, H. Kawaguchi, and Y. Nakatani, "Malcoda: Practical and stochastic security risk assessment for enterprise networks," *IEEE Transactions on Dependable and Secure Computing*, 2024.

[141] S. Maric, R. Baidar, R. Abbas, and S. Reisenfeld, "System security framework for 5g advanced/6g iot integrated terrestrial network-non-terrestrial network (tn-ntn) with ai-enabled cloud security," *arXiv preprint arXiv:2508.05707*, 2025.

[142] N. Hassan, K.-L. A. Yau, and C. Wu, "Edge computing in 5g: A review," *IEEE Access*, vol. 7, pp. 127 276–127 289, 2019. DOI: 10.1109/ACCESS.2019.2938534.

[143] R. Zhang and Q. Zhu, "Flipin: A game-theoretic cyber insurance framework for incentive-compatible cyber risk management of internet of things," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2026–2041, 2019.

[144] A. L. P. T. Bouom, J.-P. Lienou, W. E. Geh, F. F. Nelson, S. Shetty, and C. Kamhoua, "Triassetrank: Ranking vulnerabilities, exploits, and privileges for countermeasures prioritization," *IEEE Transactions on Information Forensics and Security*, 2024.

[145] H. Abie and S. Pirbhulal, "Autonomous adaptive security framework for 5g-enabled iot," *arXiv preprint arXiv:2406.03186*, 2024.

[146] E. Rodriguez et al., "A security services management architecture toward resilient 6g wireless and computing ecosystems," *IEEE access*, 2024.

[147] P. Li and J. Du, "Brand design data security and privacy protection under 6g network slicing architecture," *International Journal of Network Management*, vol. 35, no. 2, e70009, 2025.

[148] W. Li, M. Yan, and C. A. Chan, "A low-latency terminal mutual access architecture based on 5g lan cross-domain networking," in *2023 IEEE 11th International Conference on Computer Science and Network Technology (ICCSNT)*, IEEE, 2023, pp. 69–74.

[149] J. Liu, X. Wang, K. Ren, Y. Zhou, and M. Li, "Secure service function chain provisioning for task offloading in device-edge-cloud computing," *IEEE Transactions on Information Forensics and Security*, 2025.

[150] H. Fang, Z. Xiao, X. Wang, L. Xu, and L. Hanzo, "Collaborative authentication for 6g networks: An edge intelligence based autonomous approach," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 2091–2103, 2023.

[151] I. S. Forum, *Information risk assessment methodology 2 (iram2) @misc*, Nov. 2016. [Online]. Available: https://www.securityforum.org/tool/information-risk-assessmentmethodology-iram2/.

[152] F. Z. Safaeipour, J. Chakareski, and M. Hashemi, "Bayes-split-edge: Bayesian optimization for constrained collaborative inference in wireless edge systems," in *Proceedings of the Tenth ACM/IEEE Symposium on Edge Computing*, 2025, pp. 1–13.

[153] L. Alevizos and V.-T. Ta, "Threat-informed cyber resilience index: A probabilistic quantitative approach to measure defence effectiveness against cyber attacks," *arXiv preprint arXiv:2406.19374*, 2024.

[154] I. Iso, "Iec 27005: 2018—information technology—security techniques—information security risk management," *Standard. Geneva, CH: International Organization for Standardization*, 2011.

[155] M. (by Cisco), *2024 miercom sse benchmark report: Cisco secure access security service edge*, White paper, https://www.cisco.com/c/en/us/products/security/secure-access/miercom-sse-benchmark-report.html, Nov. 2024.

[156] W. Chen, C. Ju, T. Yuan, Y. Zhan, M. Zhang, and D. Wang, "Free space optical semantic communication for satellite remote sensing image transmission," *IEEE Transactions on Communications*, 2025.