Resource Management, Security, and Privacy Issues in Semantic Communications: A Survey

Dongwook Won, Geeranuch Woraphonbenjakul, Ayalneh Bitew Wondmagegn, Anh Tien Tran, Donghyun Lee, Demeke Shumeye Lakew, and Sungrae Cho

***Abstract*—Resource management, security, and privacy stand as fundamental pillars for the reliable and secure operation of efficient semantic communications (SC) system. By addressing these aspects, SC system can pave the way for efficient resource utilization, improved network efficiency, enhanced communication performance, and protection of sensitive information. In this study, we begin by presenting the background of SC and reviewing several existing studies in this field. Subsequently, we provide a comprehensive and exhaustive survey of resource management, security, and privacy in SC. We identify and highlight existing challenges and open research challenges related to resource management, security, and privacy in SC in order to spur further investigation in these areas.**

***Index Terms*—Semantic communications, Resource management, Security, Privacy**

I. Introduction

A. Background

According to the classic information theory established by Claude Shannon in 1948 [1], advances in communication systems have been driven by exploring new spectrum utilization methods and developing new coding schemes. The evolution from 1G to 5G in wireless communication has focused on achieving higher capacity, reliability, and lower latency while reducing the uncertainty related to the accurate reception of exchanged data. This relentless pursuit has led to a continuous race for wider bandwidths and higher frequency bands. However, future wireless systems like 6G must address the complex and stringent requirements of emerging applications such as the metaverse, holographic teleportation, and digital twins [2]. The interconnection of these applications will generate a staggering amount of data on the order of zettabytes. Additionally, these applications need to support massive connectivity over limited spectrum resources while requiring lower latency. This poses significant challenges to 5G systems, such as channel capacity nearing the Shannon limit [3], [4], source coding efficiency close to the Shannon information entropy/rate distortion function limit [5], and scarcity of spectrum resources [6]. Consequently, the necessity for a paradigm shift from Shannon's legacy becomes evident to accommodate the demands of next-generation communication systems, emphasizing the need for innovative approaches to meet future challenges.

The need for a paradigm shift from Shannon's legacy is evident given the limitations of current communication systems in handling the complex requirements of future applications. This shift has led to the emergence of SC. According to Shannon and Weaver's seminal work [7], communication can be divided into three levels: the technical level (Level A), the semantic level (Level B), and the effectiveness level (Level C).

- Technical Level (Level A): This level addresses the technical problem of accurately transmitting symbols from the transmitter to the receiver. It is primarily concerned with the fidelity of signal transmission, which has been the focus of traditional communication systems guided by Shannon's information theory.
- Semantic Level (Level B): This level focuses on how precisely the transmitted symbols convey the intended meaning. SC extracts content-related, task-oriented features from raw data and transmits them. This approach reduces communication resource overhead, such as bandwidth and power consumption, while ensuring the meaningful exchange of information.
- Effectiveness Level (Level C): This level considers the impact of the received information on the receiver's goal, emphasizing how effectively the meaning influences the desired outcome. Communication at this level is referred to as goal-oriented communication.

Traditional communication systems primarily operate at Level A, but the emerging demands of 6G applications necessitate a shift to Levels B and C. This paradigm shift is crucial for the development of 6G systems, which must support massive connectivity, low latency, and efficient data processing to meet the stringent requirements of next-generation applications.

B. Motivation

Traditional resource allocation schemes focus on the physical characteristics of messages, such as their size, processing capacity, and bandwidth requirements, aiming to minimize resource usage and maximize efficiency. However, SC requires resource management at the semantic level, necessitating a redefinition of performance metrics and a redesign of resource

This work was supported in part by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information Communications) and in part by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00453301) (*Corresponding author: Sungrae Cho.*)

D. Won, G. Woraphonbenjakul, A. B. Wondmagegn, A. T. Tran, D. Lee, and S. Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea (E-mail: dwwon@uclab.re.kr, geeranuch@uclab.re.kr, ayalneh@uclab.re.kr, attran@uclab.re.kr, dhlee@uclab.re.kr; srcho@cau.ac.kr).

D. S. Lakew is with the Department of Computer Science, Kombolcha Institute of Technology, Wollo University, Dessie 1145, Ethiopia (E-mail: demeke@uclab.re.kr).

allocation methods to enhance performance and efficiency from the SC perspective [8]. Resource management in SC is crucial due to its unique challenges and requirements. This includes not only allocating communication resources such as bandwidth and transmission power but also prioritizing semantic-aware transmission to ensure that more semantically significant information is sent within a given network condition. While existing literature, including surveys, briefs, and tutorials, have explored facets of SC, there is a significant research gap in the domains of resource management. Therefore, a comprehensive survey is needed to address these aspects effectively.

Distinctively, SC introduces novel security and privacy challenges compared to traditional communication systems. It is important to understand the potential vulnerabilities in different components of an SC system, such as the encoder/decoder, KB, and transmission channel, in order to propose effective countermeasures. Examining security and privacy concerns in detail will help identify risks and allow for the development of reliable techniques to enhance the protection of sensitive semantic information and user privacy. Despite SC's potential, comprehensive surveys on its security and privacy aspects are limited, highlighting the need for systematic and thorough exploration of these topics.

To bridge the gap in understanding resource management, security, and privacy in SC, this survey first constitutes a tutorial on resource management, including concepts and taxonomy. It then provides a comprehensive survey of resource management, security, and privacy, along with the redefined performance metrics used to evaluate SC network systems, by extensively reviewing the available literature. Additionally, it identifies and discusses challenges, open problems, and future research directions in these areas. By offering a thorough review of these aspects, the survey provides valuable insights and knowledge to researchers and practitioners in the field. Overall, this survey plays a crucial role in advancing the understanding and development of resource management, security, and privacy in SC.

C. Scope and Contribution

This paper mainly focuses on resource management, security, and privacy in SC to provide a comprehensive overview of the research landscape. It aims to fill the existing gap in the literature by not only covering the basics but also delving into the intricacies of resource management, security, and privacy in SC. We summarize our contributions as follows:

- **Constitute Tutorial**: This paper provides a comprehensive background tutorial on resource management, security, and privacy in SC, including concepts and an overarching framework, as detailed in Section III-A, Section IV.
- **Comprehensive Review**: Our survey offers a holistic review of the existing research on resource management, security, and privacy in SC. It covers various aspects of resource management, including semantic-aware information transmission, semantic-aware resource allocation, adaptive and optimal control of semantic compression, and joint optimization for control and allocation of resources. It also covers security and privacy issues and their countermeasures in SC. This survey contributes to the field in a broader and deeper manner compared to existing surveys. Details are discussed in Section III, IV and V.
- **In-depth Exploration**: Unlike existing surveys, our survey paper delves into the intricacies of resource management, security, and privacy in SC. It provides detailed research findings and explores the challenges and open issues in these areas. By going beyond the basics, the paper offers a deeper exploration of the research landscape in resource management.
- **Systematic Classification**: We provide a taxonomy of schemes related to resource management, security, and privacy in SC. It classifies different approaches proposed in the literature, allowing for a better understanding of the research landscape.
- **Review of Semantic Metrics**: We discuss redefined semantic metrics used in validating the performance of SC, highlighting the importance of evaluating system efficiency, effectiveness, and task-oriented performance.
- **Identification of Gaps and Future Research Directions**: Our survey identifies gaps in the existing literature and highlights open research problems. It provides new insights into both over-explored and under-explored areas, paving the way for future research directions. By doing so, the paper contributes to the advancement of knowledge in resource management, security and privacy in SC. Details are discussed in Section VI.

D. Organization

As depicted in Fig. 1, this paper systematically covers a wide spectrum of research on resource management, security, and privacy in SC. The paper is organized into several key sections. Starting with Section I, we highlight the importance and motivation for studying these aspects within the context of SC. This sets the foundation for the subsequent discussions. Moving on to Section II, we distinguish our survey from existing related surveys and tutorials, underscoring its more expansive exploration of the topics. In particular, this section underlines our survey's breadth and depth, emphasizing its comprehensive approach in capturing the nuances of the subject.

In Section III, we discuss the conceptual foundations and general workflow of SC systems, including semantic extraction and representation, semantic compression, semantic information transmission and resource management, semantic decoding, and data recovery and pragmatic function. Each phase is explained in detail, providing a clear understanding of the entire process of semantic information generation, transmission, and usage in networked systems. Building on this overview, we then present a systematic review of the literature on resource management, security, and privacy schemes proposed for SC in the research community in Section III, IV and V. Section IV provides an overview of security and privacy attacks and reviews them in detail. Additionally, in Section V, we review

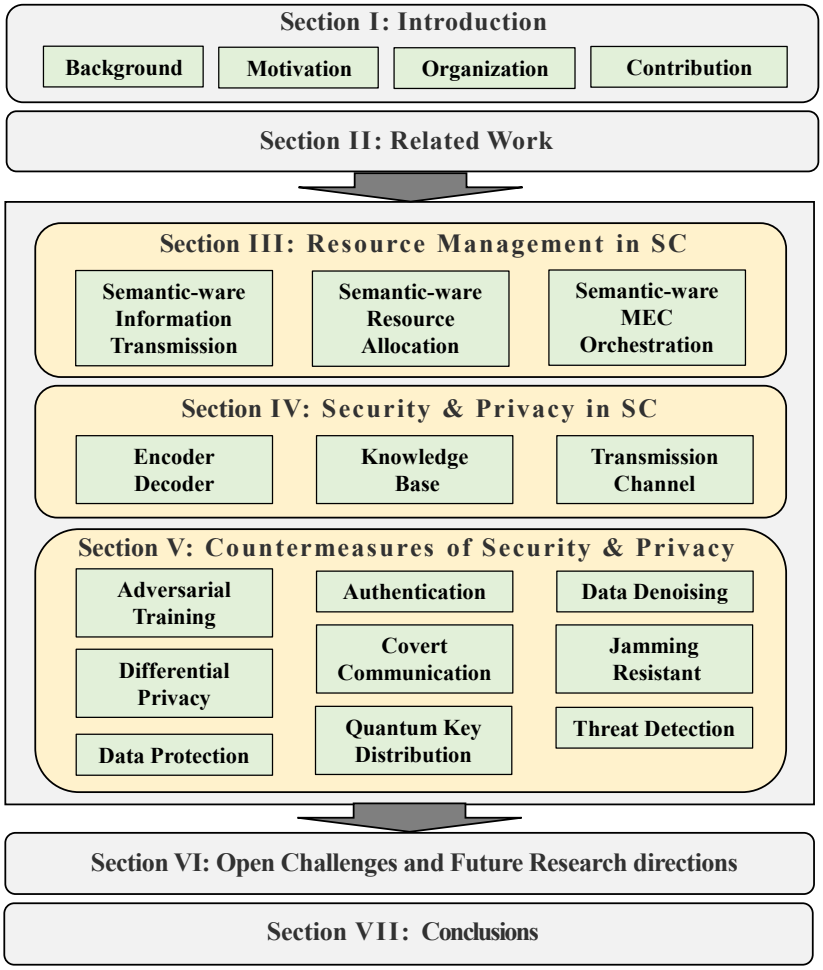


Fig. 1: Organization of Our Proposed Survey

countermeasures against these attacks. Notably, each of these sections delivers a taxonomy of the different methodologies.

In Section VI, we identify gaps in the existing literature and open research problems. It suggests several promising future research directions to advance the field further. Section VII concludes the paper. We believe that the paper contributes towards establishing a better understanding of SC and guiding optimized, secure and privacy-aware development of these systems. Common acronyms used in this survey are summarized in Tables I, II, III.

TABLE I
DEFINITIONS OF ABBREVIATIONS

Abbreviation	Definition
AD	Autonomous Driving
AI	Artificial Intelligence
AoI	Age of Information
AoII	Age of Incorrect Information
AR	Augmented Reality
ASR	Automatic Speech Recognition
BA	Bandwidth Allocation
BLEU	Bilingual Evaluation Understudy
BS	Base Station
CLM	Clinical Language Model
CNN	Convolutional Neural Network

TABLE II
DEFINITIONS OF ABBREVIATIONS

Abbreviation	Definition
CS	Charging Station
CSI	Channel State Information
DDPG	Deep Deterministic Policy Gradient
DeepJSCC	Deep Learning-based Joint Source-Channel Coding
DeepSC	Deep Learning-enabled SC
DL	Deep Learning
DNN	Deep Neural Network
DQN	Deep Q Network
DRL	Deep Reinforcement Learning
EV	Electric Vehicle
GAN	Generative Adversarial Network
H2H	Human-to-Human
HAR	Human Activity Recognition
HARQ	Hybrid Automatic Repeat Request
HSC	Hybrid Semantic Compression
IoDT	Internet of Digital Twins
IoE	Internet of Everything
IoT	Internet of Things
JSCC	Joint Source-Channel Coding
KB	Knowledge Base
KG	Knowledge Graph
LSTM	Long Short Term Memory
MEC	Multi-access Edge Computing
MI	Model Inversion
ML	Machine Learning
M2M	Machine-to-Machine

TABLE III
DEFINITIONS OF ABBREVIATIONS

Abbreviation	Definition
MSP	Metaverse Service Provider
NGNI	Next-Generation Networking Infrastructure
NFV	Network Function Virtualization
NOMA	Non-Orthogonal Multiple Access
NOMASC	Non-Orthogonal Multiple Access based SC
OMA	Orthogonal Multiple Access
PERSF-SC	Personalized Sailency Fused SC
PSNR	Peak Signal-to-Noise Ratio
QKD-SIC	QKD-Secured Semantic Information Communication
QoE	Quality of Experience
QoS	Quality of Service
QSC	Quantum SC
RNN	Recurrent Neural Network
ROI	Regions of Interest
RSU	Roadside Unit
SEC	Satellite-borne Edge Cloud
S-Rate	Semantic Transmission Rate
S-SE	Semantic Spectral Efficiency
SACCT	Soft Actor-Critic Communication Transformer
SC	Semantic Communications
SDN	Software Defined Networking
Seb	Semantic Base
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index
STM	System Throughput in Message
TOSCN	Task-Oriented SC Network
UAV	Unmanned Aerial Vehicle
UA	User Association
UE	User Equipment
URLLC	Ultra-Reliable and Low-Latency Communications
VQA	Visual Question Answering
VSO	Virtual Service Operator
VSSC	Visible, Semantic, Sample-specific, and Compatible
VR	Virtual Reality
XR	Extended Reality

II. RELATED WORK

Recent surveys on resource allocation in 5G and beyond networks have offered comprehensive insights into various strategies and challenges. Sharma *et al*. [20] and Ejaz *et al*. [21] provided taxonomies for resource allocation, with Sharma *et al*. focusing on ultra-dense networks and Ejaz *et al*. on cloud radio access networks, addressing user assignment, spectrum management, and power allocation. Studies by Manap *et al*. [22] and Agarwal *et al*. [23] delved into 5G HetNet resource management, emphasizing spectrum optimization, power allocation, interference management, and the critical role of radio resource management. They presented taxonomies for interference management and user association-resource-power allocation. Ebrahimi *et al*. [24] examined network slicing in 5G/6G, highlighting cross-domain resource management and integrated approaches for improved functionality. Lastly, Olwal *et al*. [25] discussed the technical advancements and challenges in 5G radio access network systems, focusing on multitier communication and energy-efficient management. Collectively, these surveys underscore the complexity and evolving nature of resource allocation in next-generation networks, highlighting the need for innovative solutions to meet future demands.

Several surveys and tutorials, such as those in [11]–[13], have explored various aspects of resource management in SC. In [11], Chaccour *et al*. addressed the challenges in building SC networks, noting the lack of clear definitions and technical foundations. They proposed a holistic vision for SC networks integrating artificial intelligence (AI), causal reasoning, transfer learning, and minimum description length theory, emphasizing the creation of minimal, generalizable, and efficient semantic representations of data and the importance of a facilitating semantic language. Lan *et al*. in [12] focused on efficient 6G communication through SC, classifying human-to-human (H2H) SC, human-to-machine (H2M) SC, machine-to-machine (M2M) SC, and knowledge graph (KG)-based SC, and discussed SC's potential in services like extended reality (XR), holographic communication, and all-sense communication. Qin *et al*. [13] examined SC to develop efficient SC systems across various data modalities and applications. Despite their contributions, these papers [11]–[13] lacked a comprehensive analysis of resource management and did not specifically address resource management and security.

Several brief studies [10], [26] provided various insights into new SC architecture. In [10], the authors explained the basic concepts and components of SC models, reviewed classical SC frameworks, and discussed key challenges such as semantic information extraction, knowledge modeling, and coordination, as well as data protection. They proposed an architecture based on federated edge intelligence to support semantic-aware networking, allowing users to offload computationally intensive tasks like semantic encoding and decoding to edge servers while protecting proprietary information. In [9], the authors presented an SC framework for real-time control systems like smart factory systems, considering control signals as semantic information. They provided a multi-granularity definition of semantic information across different communication system levels and illustrated SC through a system monitoring two sources affecting a robotic object to create a digital twin. By employing semantics-empowered sampling and communication policies, the study demonstrated significant reductions in reconstruction error, actuation error costs, and uninformative sample generation. The work in [18] highlighted the need for joint orchestration of C4 resources in edge computing. C4 stands for communication, computation, caching, and control, and these resources are tightly coupled in the design of multi-access edge computing (MEC) systems. Zhilin Lu *et al*. [19] conducted a survey on the history, theories, metrics, and datasets related to semantic information transmission and adaptive control of the semantic compression ratio.

In [14], Yang *et al*. provided a comprehensive survey for the implementation of SC in 6G, thoroughly reviewing existing studies and discussing 6G applications in potential SC-empowered network architecture. They delved into the fundamental concepts of SC by discussing semantic representation techniques, such as word embeddings and KG, which enable machines to understand and reason about the meaning of information. Furthermore, Yang *et al*. discussed the trade-off between SC performance and data security. They

TABLE IV
QUALITATIVE COMPARISON OF EXISTING SC WORKS WITH THE PROPOSED SURVEY

Ref.	Transceiver Design	Adaptive Control	Resource Allocation	MEC Orchestration	Security & Privacy
[9]	X	X	△	X	X
[10]	X	△	△	X	X
[11]	△	X	△	X	X
[12]	O	△	△	X	△
[13]	△	△	△	X	△
[14]	O	X	△	△	△
[15]	O	X	△	△	O
[16]	X	X	X	X	O
[17]	X	X	X	X	O
[18]	△	X	X	△	X
[19]	O	O	X	X	X
Our Survey	O	O	O	O	O

Notation: In the table, Ref. means Reference, "O" means comprehensively covered, "△" means partially discussed, "X" means not discussed.

mentioned that SC can be seen as a potential method for secure communications, as it requires only partial data to be transmitted, and the decoding of sensitive information relies on the receiver's background knowledge. However, it also led to considering the trade-off between computational resource overhead and data security. The process of encoding semantic information and the subsequent decoding by the receiver, especially when complex algorithms are involved, can be resource-intensive. This led to an important consideration: the trade-off between the computational resources required and the level of data security achieved. Yang *et al*. also provided a tutorial-style overview of resource management concepts, but its scope was somewhat limited and fell short of delving into the complexities and nuances of the subject matter. Similarly, the work in [15] touched upon algorithmic developments in resource allocation and the security and privacy aspects of SC, but its coverage remains more introductory.

However, the aforementioned papers [9], [10], [14], [15], [18], [19] lack detailed research findings, particularly in the areas of resource management, security, and privacy in SC. This gap in the literature highlights the need for a more systematic and comprehensive survey that not only covers the basics but also delves into the intricacies of resource management, security, and privacy in SC. This is precisely the motivation behind our proposed survey, which aims to fill this critical gap in the existing research.

Recent surveys on security problems targeting 5G and 6G emerging technologies offer critical insights. These studies have focused on network slicing in MEC [27], [28], massive MIMO [29], D2D communication [30], and SDN/NFV [31], [32], [33]. Wang *et al*. [34] reviewed various security challenges and issues, while Schmittner *et al*. [35] and Alturfi *et al*. [36] discussed several security attacks targeting 5G and beyond networks. However, these papers [27]–[36] have ignored or under-analyzed SC. Consequently, there is a pressing need for comprehensive security studies in SC.

Some papers [16], [17] have focused on security and privacy. The authors in [16] discussed SC and proposed four methods to represent semantic information: semantic entity, KG, probability graph, and probability distribution, highlighting their advantages and challenges. They also addressed SC security aspects, including information security and machine learning (ML) security, discussing potential threats and countermeasures in SC networks. In [17], Du *et al*. provided guidelines for designing secure SC systems in real-world wireless networks, emphasizing the next generation of wireless networking and the transition to the Semantic IoT. The authors revisited classical communication security techniques from the perspective of semantic networks and discussed the novel attack and defense methods introduced by SC techniques, along with two new performance indicators. Both papers [16], [17] offered valuable insights into SC and its security implications but shared a common shortcoming: the lack of discussion on resource management, highlighting a clear gap in the literature for future research to address.

To the best of our knowledge, our paper is the first addressing resource management, security, and privacy issues together in SC comprehensively. We provide a detailed review of several studies in this field, discussing their challenges, advantages, and shortcomings. This paper also presents a taxonomy of schemes related to resource management, security, and privacy, along with the metrics used to evaluate the network system of SC. Additionally, it highlights open issues and research directions, emphasizing the relevance of the topic. To summarize, the state-of-the-art surveys on SC are provided in Table IV.

III. RESOURCE MANAGEMENT IN SC

A. General SC Framework

The workflow of general SC framework is depicted in Fig. 2. It consists of six phases: 1) Semantic extraction and representation, 2) Semantic compression, 3) Semantic information transmission and resource management, 4) Semantic decoding, 5) Data recovery and pragmatic function, and 6) Performance evaluation.

1) Semantic Extraction and Representation: As stated in [37], information is a commodity that yields knowledge, and the value of a signal lies in what we can learn from it. Unlike

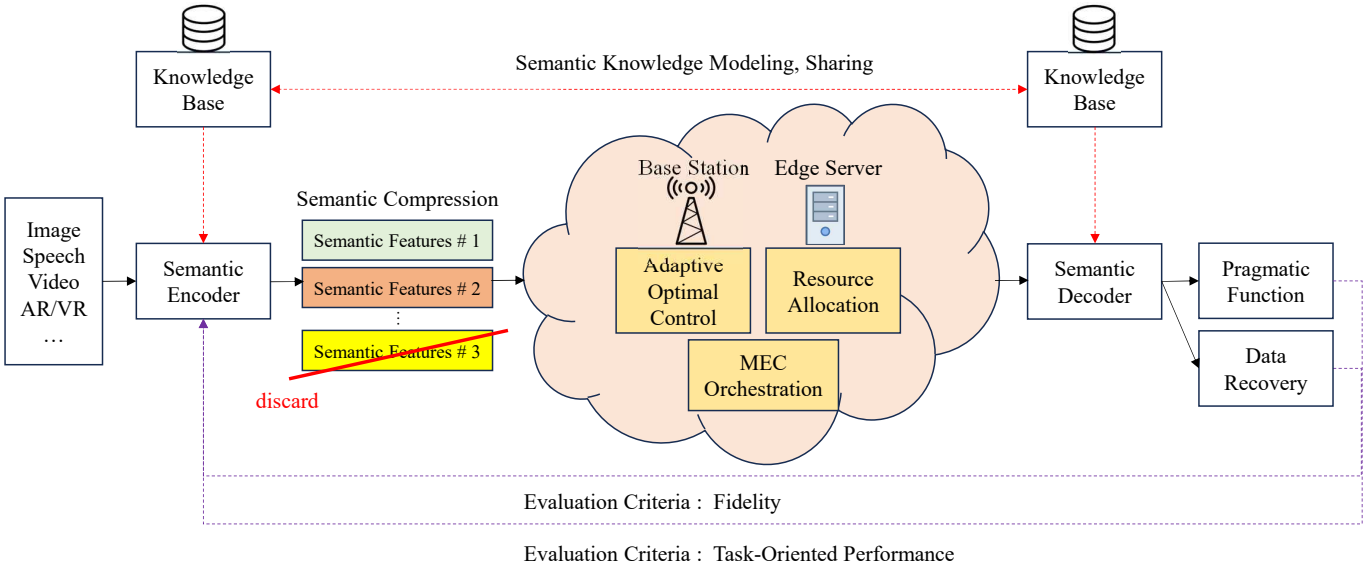


Fig. 2: Comprehensive Overview for Understanding Resource Management in SC

Shannon information, which focuses on how often a message occurs regardless of its content or relevance to the task of the receiver, semantic information focuses on how often a message occurs that can be considered 'true' as agreed upon by the transmitter and the receiver. If a message is more commonly true, it contains less semantic information [38]. By transmitting only semantic information, SC allows the generation of more knowledge with less data. To achieve SC, the transmitter and receiver first define a semantic distortion (or loss) function, which they then minimize by training a deep neural network (DNN) of the semantic encoder and decoder. Through this training, the semantic encoder and decoder are able to successfully extract semantic features related to 'true' from the raw data, allowing them to transmit semantic information. Consequently, a crucial research question arises: How effectively can semantic encoder and decoder extract and represent these semantic features? This topic is explored in detail in the related works discussed in Section III-B.

Knowledge representation plays a pivotal role in semantic information extraction, involving the representation of inherent semantic features within raw data. One common representation technique is the KG, which is particularly powerful as it can represent the logical relationships among semantic features. By organizing semantic features into a structured format, KG facilitates further processing and analysis. Another technique, the Bayesian network model, captures conditional dependencies among semantic features based on expert knowledge and data analysis. For instance, in image analysis, semantic features like 'sky' and 'grass' are identified through object detection algorithms, aiding in scene classification [39]. These features are subsequently analyzed in depth using a Bayesian network, which elucidates their interactions with other image elements. Throughout this process, the knowledge base (KB) is integral, systematically storing and managing this structured semantic data to support efficient retrieval and reasoning, thereby aiding complex decision-making and problem-solving across various applications.

2) Semantic Compression: In semantic compression phase, the goal is to compress the correlated semantic features while preserving the essential semantic information. To achieve this, important semantic features are preserved, while less important ones are discarded. This process is carried out using a suitable compression algorithm. The compressed semantic representation is organized into a sequence of semantic features, which is commonly referred to as a semantic stream when prepared for transmission. At that time, semantic compression ratio is important. A higher semantic compression ratio results in greater semantic distortion, negatively affecting task performance. Conversely, a lower semantic compression ratio preserves more detail but increases the communication load. This balance is formulated based on the rate-distortion trade-off, ensuring an optimal compromise between data fidelity and compression efficiency. Additionally, a mechanism to control the rate according to the given network situation is required to adapt dynamically.

3) Semantic-aware Information Transmission and Resource Management: The semantic stream is then transmitted, where it may be affected by noise and interference. The network system also dynamically controls the compression ratio based on current network conditions. Resource management, such as the allocation of bandwidth and power, is also performed at this stage. However, unlike traditional communication systems that define system performance metrics at the message level, in SC, performance metrics need to be defined at the semantic level.

In message-level resource management, communication occurs on a per-message basis, with each message potentially containing specific resource requirements. These resource requirements are related to message characteristics such as message size, processing capacity, bandwidth, and so on. On the other hand, at the semantic level, the focus is on the meaning and purpose of the communication. Semantic-level

resource management considers the intent, and meaning of the communication. This involves considering the content of the messages, semantic related requirements, and other factors to make semantic-aware resource allocation decision. Therefore, in SC, it is need to redefine metrics at the semantic level and perform resource management accordingly. This allows for optimization of performance, reliability, efficiency, and effectiveness in the SC system.

4) Semantic Decoding: Subsequently, the distorted semantic information is fed into the semantic decoder, which generates an output using the existing KB [40]. The decoding process leverages advanced deep learning (DL) technologies, such as Transformers and auto-encoders, which are powerful tools for handling KB. The main goal of SC is to ensure that the receiver understands the meaning of the data. To achieve this, both the semantic encoder and decoder are jointly trained to meet the semantic metric or user pragmatic task performance.

5) Data Recovery and Pragmatic Function: Data recovery refers to the process of reconstructing data that has been compressed back to its original or near-original form. In SC, this involves using decoded semantic features to recreate the original source data as closely as possible, aiming for high fidelity in content, quality, and meaning. In H2H scenarios, high fidelity in data recovery is crucial. Unlike M2M and H2M interactions, where operational correctness and efficiency are paramount, H2H interactions require preserving the original meaning for tasks like natural language processing or image recognition. Therefore, in H2H scenarios, it is important to design transceivers that consider this trade-off between fidelity and efficiency.

The pragmatic function models the user tasks, such as image classification, segmentation, and reconstruction, etc. It takes the decoded semantic information or reconstructed data as its input to execute these tasks depend on communication scenario and then utilized by the pragmatic function to carry out specific tasks that the user or system requires. For example, if the user task involves image classification, the pragmatic function would use the decoded semantic features to identify and categorize the objects within an image. The effectiveness of the pragmatic function depends on the quality of the decoded semantic information. Accurate semantic decoding enables the pragmatic function to perform tasks more effectively and reliably, enhancing the overall user experience.

6) Performance Evaluation: The effectiveness of the pragmatic function is assessed using a variety of metrics tailored to the specific task at hand, broadly categorized into task-oriented performance, fidelity, and system-oriented performance. Task-oriented performance metrics vary based on the specific task, such as intersection over union for image segmentation accuracy and classification accuracy for image classification tasks. The fidelity measures how closely reconstructed or transmitted data reflect the semantic meaning or original data, using metrics like peak signal-to-noise ratio (PSNR) for images and bilingual evaluation understudy (BLEU) for text. At the semantic level, text semantic reconstruction is evaluated using semantic similarity to assess the quality of reconstruction. System-oriented performance focuses on the efficiency and reliability of the communication system, incorporating energy and spectral efficiency, Quality of Service (QoS) metrics like transmission rate, delay, and throughput, as well as Quality of Experience (QoE) for user satisfaction. Additionally, robustness, or the system's ability to perform well under challenging network conditions, is evaluated.

Based on the general framework presented above, this survey aims to serve as a comprehensive overview for understanding resource management in SC. As depicted in Fig. 3, the recent literature review is divided into four major sections, each focusing on a unique dimension of the complex issue of resource management.

First, Section III-B emphasizes the design of semantic-aware information transmission systems to enhance task-oriented performance or fidelity. It covers extracting semantic features, optimizing and adapting compression ratios, and joint resource allocation management in SC systems. This section explores methods to reduce the size of transmitted data while preserving their original meaning, focusing on optimizing the semantic compression ratio and resource allocation. It is structured into three main categories of research. First, Section III-B1 focuses on extracting essential semantic features from large datasets to design semantic-aware transciver that improve task and system performance or fidelity. These studies analyzed the rate-semantic distortion trade-off. Second, Section III-B2 examines the determination of optimal semantic compression ratios that adapt to changing network conditions by analyzing the relationship between compression ratios and system-oriented or task-oriented performance metrics, building upon the foundational systems discussed earlier. Finally, Section III-B3 goes further by not only controlling the amount of transmitted semantic information but also optimizing resource allocation, leveraging the relationship between semantic compression ratios and performance metrics for joint optimization of semantic compression control and resource allocation.

Second, Section III-C introduces wireless resource management from a networking perspective. Traditional metrics based on Shannon's theory are no longer applicable in SC, which emphasizes the meaning of delivered messages over transmitted bits. Therefore, new units like semantic units or semantic bases (Sebs) are introduced. Consequently, metrics such as transmission rate and spectral efficiency are redefined as semantic transmission rate and semantic spectral efficiency using Seb. This allows SC system performance to be measured and optimized more effectively at the semantic level. Existing papers addressing this challenge are explored in Section III-C1.

The second another challenge focuses on optimizing semnatic-aware resource allocation utilizing newly defined semantic metrics. Optimal resource allocation strategies vary based on criteria such as QoS, QoE, energy efficiency, and spectral efficiency. Section III-C2 will review papers optimizing resource management from a QoS perspective, focusing on transmission rate, delay, and throughput. Section III-C3 will examine papers that address energy efficiency optimization, while Section III-C4 will explore papers focusing on spectral efficiency.

Lastly, Section III-D discusses semantic-aware resource

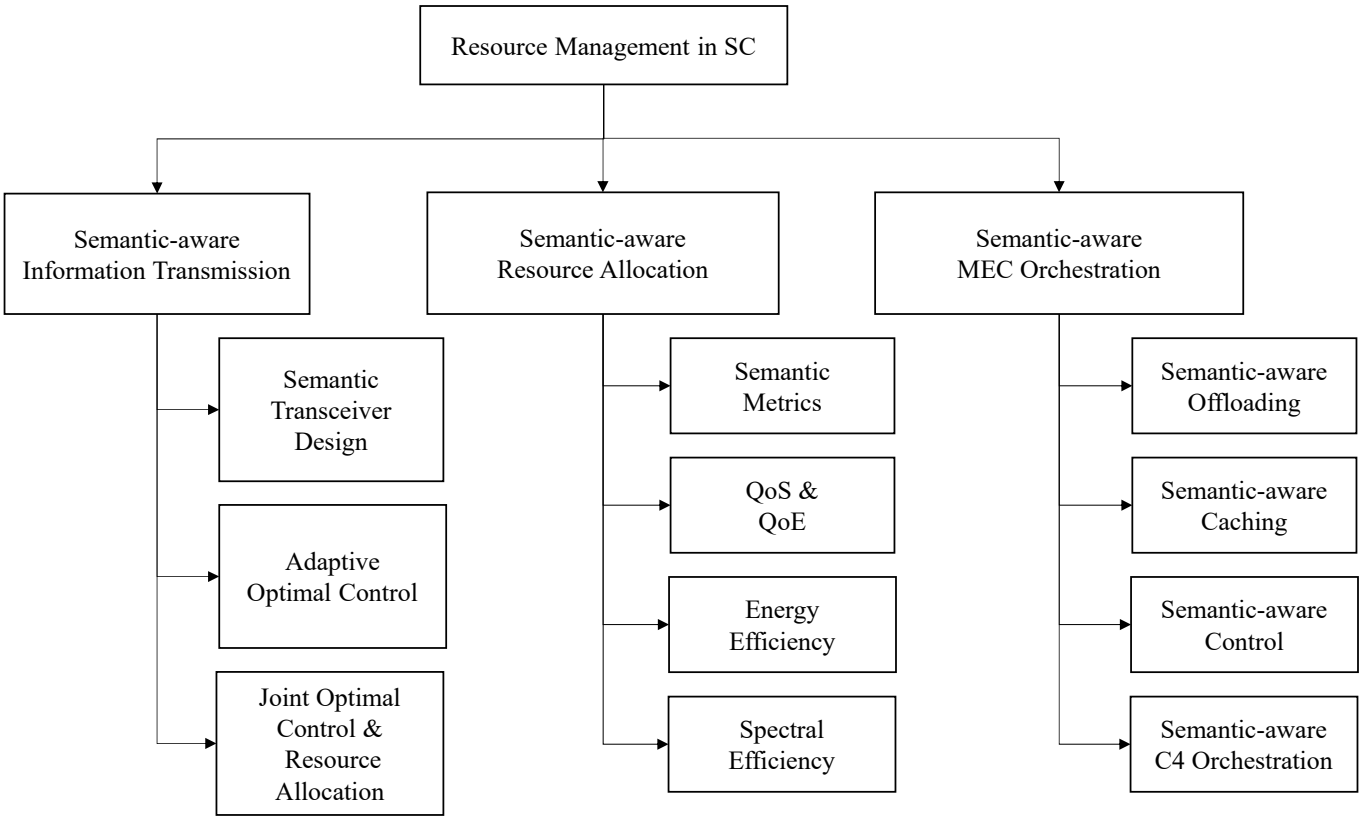


Fig. 3: Taxonomy of Resource Management in SC

management techniques in MEC environment. Section III-D1 covers offloading in SC, Section III-D2 addresses caching, Section III-D3 discusses control, and Section III-D4 explores the integration and coordination of communication, computation, caching, and control (C4) resources in SC networks. This integration is essential for efficient data handling and processing in resource-constrained MEC. By the end of this survey, the readers will have a comprehensive understanding of SC resource management.

B. Semantic-aware Information Transmission

1) Semantic-aware Transceiver Design: In SC systems, a lot of research has been dedicated to designing and optimizing systems that maximize task-oriented performance. Deep learning has been integrated into various information-theoretic and link-level systematic modeling approaches, leading to a diverse range of variants depending on the source type. Among these, the deep learning-enabled semantic communication (DeepSC) system [41] and its several variants stand out. These variants cater to different types of data sources, including text [41]–[44], visual [45]–[49], speech [50]–[52], and multimodal [53]–[55].

For text SC systems, semantic sentence transmission is critical, with many state-of-the-art methods being proposed to enhance performance and efficiency. Farsad *et al*. [56] pioneered the initial deep learning-based joint source-channel coding (DeepJSCC) for text transmission, utilizing a recurrent neural network (RNN) and a fully-connected neural network to encode text sentences into fixed-length bit streams over simple channel environments. This system directly recovers text without separate channel and source decoding, demonstrating superiority under high bit drop rates. Building on this foundation, Xie *et al*. [41], [42] developed DeepSC, a SC system based on the Transformer model, which introduced the concepts of semantic information and semantic error at the sentence level. Compared to traditional approaches, DeepSC is more robust to channel variations and achieves better performance in source recovery, particularly in low signal-to-noise ratio (SNR) regimes. The DeepSC system includes a transmitter with a semantic encoder to extract semantic features and a channel encoder to generate transmission symbols, while the receiver comprises a channel decoder for symbol detection and a semantic decoder for text estimation. Trained using a loss function that considers sentence similarity and mutual information, DeepSC was evaluated using the BLEU score and found to outperform traditional methods and other deep learning-based networks, especially in low SNR conditions. Xie *et al*. further extended DeepSC to the Internet of Things (IoT) with a distributed SC system for capacity-limited networks, addressing practical issues such as channel impact, quantization, and network compression [43]. Additionally, Zhou *et al*. [44] proposed a cognitive SC framework leveraging KG to surpass the Shannon limit. This framework utilizes triples as semantic symbols, enabling error correction at the symbolic level and enhancing communication reliability while reducing data compression rates.

For image SC systems, Bourtsoulatze *et al*. [57] conducted pioneering work by jointly optimizing semantic and channel coding. They recognized the capability of DNNs for image compression and channel noise estimation, proposing an auto-encoder based DeepJSCC model to transmit images over a wireless channel. Unlike traditional image compression, which transforms images into bits, their DeepJSCC approach maps pixel values to complex-valued channel symbols. However, stochastic noise in wireless channels posed significant challenges, leading to performance degradation. To address this, Kurka *et al*. [58] introduced DeepJSCC with feedback, where the received signal is fed back to the transmitter to mitigate channel noise effects, improving image reconstruction quality. Building on these advancements, Xu *et al*. [59] proposed attention mechanism-based DeepJSCC, which enhances noise tolerance by dynamically adjusting bit allocation between source and channel encoders based on current SNR conditions. The key innovation of attention-based DeepJSCC is its use of a channel-wise soft attention network for adaptive compression, providing robustness across various SNR conditions. In contrast to previous works that focus on physical noise, the authors in [45] addressed both channel and semantic noise, proposing adversarial training with weight perturbation and a masked vector quantized-variational autoencoder to enhance system robustness. Additionally, Zhang *et al*. [46] discussed the need for domain adaptation in practical scenarios. They proposed a SC system for image transmission focused on raw image reconstruction and task performance enhancement, using PSNR and accuracy metrics. They identified the out-of-distribution problem and proposed a domain adaptation approach, integrating a data adaptation network to maintain high performance in varying datasets, thereby optimizing the balance between task performance and image reconstruction.

In addition to image transmission systems, research has also focused on other types of visual data, such as point cloud data [48] and video [49]. Varischio *et al*. [48] proposed a hybrid semantic compression (HSC) pipeline algorithm for LiDAR point clouds to achieve real-time transmissions in autonomous driving scenarios. LiDAR data, which generate large volumes of point clouds carrying geometry and attribute information for object detection, localization, and recognition, need efficient compression for real-time communication within limited bandwidth channels. The HSC pipeline combines Google's Draco software [60] for compression with RangeNet++ [61], for semantic segmentation, classifying data points and identifying valuable objects like pedestrians and vehicles. The pipeline supports three transmission levels, each progressively removing less critical elements to prioritize valuable data, achieving up to 700 times compression with tolerable accuracy degradation. Similarly, Wang *et al*. [49] addressed data reduction and automotive camera video compression for assisted and automated driving functions, proposing a semantic-aware video compression framework. This method separates each video frame into regions of interest (ROI) and non-ROI, applying different compression ratios to maintain the quality of critical navigation information. Future research could explore more advanced codecs, improve semantic segmentation accuracy, and refine the definition of ROI and non-ROI for various applications, aiming for a comprehensive evaluation of image quality in automotive contexts.

Recent advances in SC for speech transmission have demonstrated significant improvements in both efficiency and accuracy. Weng *et al*. [50] developed an attention-based SC system utilizing CNNs to compress speech spectra, treating each frame as an image, and leveraging the squeeze-and-excitation network for superior performance. Tong *et al*. [51] enhanced accuracy through federated learning by training across multiple devices. Additionally, Han *et al*. [52] proposed a two-stage training scheme to accelerate training times. These methods have proven to be more efficient than traditional communication systems, reducing character-error-rate and word-error-rate, and performing robustly across various channel conditions. Consequently, SC has emerged as a promising solution for transmitting critical semantic information in bandwidth-limited environments.

The aforementioned SC systems have exhibited satisfactory performance in certain scenarios, primarily focusing on single-modal data. However, the evolution of SC systems has highlighted the need to support multimodal data from different users efficiently [62]. Initial research by Xie *et al*. [63] introduced MU-DeepSC, a task-oriented multi-user SC system designed to enhance the accuracy of visual question answering (VQA) tasks. This system uses Long Short Term Memory (LSTM) for text transmission and Convolutional Neural Network (CNN) for image transmission, demonstrating robustness to channel variations, particularly in low SNR environment. Building on this, Xie *et al*. [53] expanded their framework to unify the semantic encoding structure for both image and text transmitters using Transformer models. This extension addresses inter-user interference and diverse data distribution fusion, making it suitable for various autonomous applications in daily life and industry. Further advancing this field, Li *et al*. [54] introduced a cross-modal SC paradigm designed to overcome the polysemy and ambiguity issues inherent in multimodal services. Their approach includes a cross-modal knowledge graph and advanced semantic encoder and decoder, ensuring high reliability and precise recovery of multimodal signals. These advancements collectively underscore the importance of developing robust multimodal SC systems to support diverse applications effectively. However, the need for a model that can handle multiple tasks without separate retraining remained. Zhang *et al*. [55] addressed this by proposing U-DeepSC, a unified SC framework capable of handling multiple tasks with a fixed model. U-DeepSC employs a vector quantized variational mechanism for discrete feature representation and a digital modulation module, significantly reducing transmission overhead and enhancing task performance.

2) Adaptive Optimal Control of Semantic Compression: The aforementioned works in the Section III-B2 discussed the impact of compression ratio on system performance and incorporated it into the system design and implementation. However, they did not address optimizing the compression ratio based on network conditions, nor did they include adaptive control mechanisms for varying network conditions. In this section, we will explore studies that consider the optimal

and adaptive control of compression efficiency, particularly in the context of network conditions, to enhance network system efficiency and improve task performance.

Conventional text-based SC systems [41], [43], [44], [56] primarily rely on fixed-length codes for sentences, which are unsuitable for sentences of varying lengths and result in inefficiencies under varying meanings and SNR conditions. To address this, Rao *et al*. [64] proposed a variable-length code based on LSTM, which performs better than the fixed-length approach in handling longer sentences. Additionally, Sana *et al*. [65] introduced a semantic adaptive mechanism that dynamically optimizes the compression ratio per semantic message, ensuring semantic accuracy by carefully evaluating the balance between semantic compression and semantic fidelity. Further enhancements can be achieved by introducing state-of-the-art semantic technologies, such as the universal Transformer [66], which can adapt to different channel conditions but still lacks the flexibility to dynamically change code lengths. Jiang *et al*. [66] tackled this limitation by combining semantic coding with Reed-Solomon channel coding and hybrid automatic repeat request (HARQ), leveraging the strengths of both semantic and conventional coding methods. This approach significantly reduces the required number of bits for semantic sentence transmission and the sentence error rate. To further address inefficiencies and lack of scalability, Zhou *et al*. [67] proposed an adaptive method using multi-bit length selection and a progressive semantic HARQ scheme with incremental knowledge to reduce communication costs and semantic errors. This method employs a policy network to decide the appropriate coding rate and introduces a specific denoiser to reduce semantic errors during transmission.

Zhang *et al*. [47] proposed an adaptive control mechanism to improve the efficiency of wireless image transmission in SC. They proposed a predictive and adaptive deep coding framework that adjusts the code rate based on channel SNRs and image contents, enhancing image quality and reducing transmission errors. Additionally, Zhu *et al*. [68] proposed AITransfer, a semantic-aware transmission method for volumetric video, which adapts compression ratios to dynamic network conditions, optimizing visual quality and transmission efficiency based on real-time bandwidth availability.

3) Joint Optimization for Control and Allocation of Resource: While previous works have significantly contributed to designing SC systems and controlling the amount of transmitted information, they did not address resource allocation in specific network conditions and requirements. Comprehensive approaches considering the joint optimization of semantic compression and resource allocation are required. Yan *et al*. [8] proposed a semantic spectral efficiency (S-SE) model to measure communication efficiency from a semantic perspective, aiming to maximize overall S-SE by optimizing the transmission volume of semantic information. Similarly, Yang *et al*. [74] addressed resource allocation and semantic information extraction in energy-efficient SC with rate splitting, formulating an optimization problem to minimize total communication and computation energy consumption while satisfying various constraints.

Liu *et al*. [70] proposed a task-oriented communication architecture for multi-user SC systems, introducing adaptable semantic compression to optimize compression ratio and resource allocation, thereby maximizing task success probability and reducing data size by up to 80%. Chi *et al*. [69] proposed a scheme for jointly optimizing compression ratio, power allocation, and resource block assignment to maximize user content reception. They demonstrated the superiority of an adaptive decision method over a fixed method for compression ratio in minimizing power allocation while meeting latency constraints.

Further advancements include Binucci *et al*. [71] introducing a comprehensive framework that leverages Lyapunov stochastic optimization to dynamically optimize various resources in edge learning environments. Their approach aims to balance communication, computation, and encoder-classifier resources effectively, thus enhancing the overall efficiency of edge learning. This method addresses the complex challenge of resource management in dynamic and distributed settings, ensuring that resources are allocated in a manner that maximizes learning efficiency while minimizing latency and energy consumption. Similarly, Yan *et al*. [73] conducted an in-depth investigation into resource allocation strategies designed to maximize the Quality of Experience (QoE) in semantic communication networks. They formulated an optimization problem that integrates semantic symbol transmission, channel assignment, and power allocation, ultimately aiming to enhance user satisfaction by optimizing the overall communication process. Additionally, Zhang *et al*. [76] employed deep reinforcement learning (DRL) techniques for resource management in task-oriented semantic communication networks (TOSCNs). Their approach dynamically adjusts resource allocation based on the current state of the system, effectively balancing the quantity of data packets and the accuracy of tasks, thereby improving the adaptability and performance of semantic communication systems in real-time scenarios.

Building on these developments, Binucci *et al*. [72] proposed a novel resource allocation approach specifically tailored for multi-user environments. This approach utilizes Lyapunov optimization to manage and allocate resources efficiently among multiple users, addressing the unique challenges posed by multi-user settings. Meanwhile, Li *et al*. [75] explored the use of non-orthogonal multiple access (NOMA) in semantic communication systems, demonstrating its superiority over traditional orthogonal multiple access (OMA) schemes. Their findings highlight significant improvements in semantic transmission rates and power efficiency, suggesting that NOMA-enabled SC could be a promising direction for future research.

The studies discussed in Section III-B underscored the significance of new semantic-aware information transmission system design for maximizing fidelity, task-oriented and system oriented performance. The research and development reflect the dedication of the academic community to pioneering advancements in this field. The trajectory of these investigations signals a promising future for SC, ensuring its central role in next-generation communication systems. A summary of existing works of semantic-aware information transmission is presented in Table V.

TABLE V
SUMMARY OF SEMANTIC-AWARE INFORMATION TRANSMISSION

Ref.	Transceiver Design	Optimal Control	Resource Allocation	Performance Metrics			Data Type
				User Task	Fidelity	System	
[41]	O	X	X	X	O	X	Text
[44]	O	X	X	X	O	X	Text
[42]	O	X	X	X	O	X	Text
[43]	O	X	X	X	O	X	Text
[44]	O	X	X	X	O	X	Text
[59]	O	X	X	X	O	X	Visual
[48]	O	X	X	X	O	X	Visual
[49]	O	X	X	O	O	X	Visual
[45]	O	X	X	O	O	X	Visual
[46]	O	X	X	O	O	X	Visual
[51]	O	X	X	X	O	X	Speech
[50]	O	X	X	O	O	X	Speech
[52]	O	X	X	O	O	X	Speech
[54]	O	X	X	X	O	X	Multimodal
[53]	O	X	X	O	O	X	Multimodal
[55]	O	X	X	O	O	X	Multimodal
[64]	O	O	X	X	O	X	Text
[66]	O	O	X	X	O	X	Text
[65]	O	O	X	X	O	O	Text
[67]	O	O	X	X	O	O	Text
[68]	O	O	X	X	O	O	Visual
[47]	O	O	X	O	O	X	Visual
[8]	X	O	O	X	X	O	Text
[69]	X	O	O	X	X	O	Text
[70]	X	O	O	X	X	O	Visual
[71]	X	O	O	X	X	O	Visual
[72]	X	O	O	O	X	O	Visual
[73]	X	O	O	X	X	O	Multimodal
[74]	X	O	O	X	X	O	Multimodal
[75]	O	O	O	X	O	O	Multimodal
[76]	O	O	O	O	X	O	Visual

C. Semantic-aware Resource Allocation

1) Metrics for Semantic-aware Resource Allocation: The optimization of usually scarce resources for optimality and efficiency across one or more networks—across wireless or optical SC networks—is one of the key problems facing classical SC systems. Semantic metrics for resource allocation are therefore crucial to optimize several types of classical SC systems. The following semantic metrics for resource allocation are largely relevant to optimizing SC systems: the metric of S-Rate, S-SE, QoE, and system throughput in messages (STM). We present these semantic-related performance metrics below, beginning with STM.

The paper [77] proposed a new metric, System Throughput in Messages (STM). The STM aimed to measure the sum of the message rates delivered to all mobile users within a time unit. Herein, the unit message, in a communication sense, indicated a complete piece of information successfully transmitted. For instance, an entire text sentence ending with a period in text communication or a voice signal completely sent out in speech communication could be regarded as a message. With that in mind, the aforementioned message rate was thus interpreted as the number of messages conveyed or processed per time unit (*msg/s*), with reference to the bit-rate definition (*bit/s*). For all $i \in U$, let $S_i(\cdot)$ denote a universal bit-to-message transformation function of MU_i under given channel conditions, which converts bit rate to message rate in a physical sense. Note that the manifestation of $S_i(\cdot)$ should be solely from the users' side, jointly determined by adopted semantic models, associated KBs, and received message properties. Then, in view of the bit rate b_{ij}, we can take advantage of $S_i(\cdot)$ to naturally define the message rate by ξ_{ij}, where $\xi_{ij} = S_i(b_{ij})$, for all $(i, j) \in U \times B$. Thus,

$$T_M = \sum_{i \in U} \sum_{j \in B} x_{ij} \xi_{ij} \tag{1}$$

Here, T_M is the STM that could well represent the network performance from a semantic perspective.

In [8], the authors proposed a new performance metric for text-based SC, which was S-Rate. S-Rate represented the

amount of semantic information I effectively transmitted per second. It was measured in semantic units per second (suts/s). The semantic unit represented the basic unit of semantic information. Specifically, the S-Rate of the n-th user over the m-th channel with bandwidth W could be expressed as:

$$\Gamma_{n,m} = \frac{WI}{k_n L}\xi_{n,m} \tag{2}$$

where $\xi_{n,m}$ signifies the semantic similarity of the n-th user over the m-th channel and relies on the neural network structure of the system. $k_n L$ denotes the expected number of semantic symbols per sentence for the n-th user. Therefore, this formula ties together the channel's capacity, the effectively transmitted semantic information, and the efficiency of the transmission. In [8], the authors also proposed S-SE based on S-Rate. The S-SE referred to the rate at which semantic information could be successfully transmitted over a unit of bandwidth. It was measured in semantic units per second per Hertz (suts/s/Hz). The corresponding S-SE could be expressed as:

$$\text{S-SE}(\Phi_{n,m}) = \frac{\Gamma_{n,m}}{W} \tag{3}$$

However, evaluating the performance of S-Rate imposed challenges. The main metric used was the similarity between the original and received data. Formulating this similarity function for theoretical analysis was very complex, and there were no closed-form expressions provided for the sentence semantic similarity function. Therefore, in [8], the value of similarity could only be obtained via experiments on DeepSC [41].

To address this challenge, the authors in [78] turned to a data regression method. They observed that the semantic similarity, when plotted against certain parameters, exhibited an 'S' shape and remained confined between two values. This observation was key as it hinted at the potential use of the generalized logistic function to approximate this semantic similarity. By leveraging this function, they could effectively model the behavior of the semantic similarity with respect to the parameters in question. The significance of this approximation cannot be understated. With a tractable function in hand, the authors could dive deeper into the intricacies of SC. This paved the way for more in-depth theoretical investigations, allowing for the optimization of systems that prioritize the transmission of semantics over mere data. In essence, the S-Rate approximation provides a foundational tool for the design and analysis of next-generation communication systems, emphasizing the importance of meaning in transmitted data.

In [73], Yan *et al*. discussed the performance metric and defined it as a QoE model. The QoE specifically consists of two components, semantic rate and semantic similarity score, corresponding to user quality of service and user task performance, respectively. Yan *et al*. focused on two types of intelligent tasks, including a single-modal task and a bimodal task. Assume that N_{Bi}^b is bimodal user pairs of q-th user group and U_q^b is the set of all user groups in the b-th cell. The semantic rate was defined as the amount of semantic information emitted to the transmission medium per second, measured in *suts/s*. The semantic rate of the single-modal user $u \in U_q^b$, where $q > N_{Bi}^b$, was given by:

$$\phi_u = \frac{\tilde{H}_{Si}}{k_u/W} \tag{4}$$

where $\tilde{H}_{Si}$ represents the DeepSC [41] based approximate semantic entropy.

The semantic rates of bimodal users u_t and u_i, where $u_t, u_i \in U_q^b$ and $q \leq N_{Bi}^b$, were expressed as:

$$\phi_{u_t} = \frac{\tilde{H}_{Bi,t}}{k_{u_t}/W} \tag{5}$$

$$\phi_{u_i} = \frac{\tilde{H}_{Bi,i}}{k_{u_i}/W} \tag{6}$$

respectively, where $\tilde{H}_{Bi,t}$ and $\tilde{H}_{Bi,i}$ represented the DeepVQA [79] based approximate semantic entropy for text transmission user and image transmission user, respectively.

The QoE model was formulated using the semantic rate and semantic accuracy. Specifically, the QoE of the q-th user group in the b-th cell, denoted as QoE_q^b, is given by:

$$QoE_q^b = \sum_{u \in U_q^b} w_u G_u^R + (1 - w_u) G_u^A \tag{7}$$

where w_u and $1 - w_u$ represented the weights of the semantic rate and semantic accuracy for user u, respectively. Additionally, G_u^R and G_u^A denoted the scores of semantic rate and semantic accuracy for user u, respectively. Based on (7), a semantic-aware resource allocation method was investigated that maximized the QoE by optimizing the number of transmitted semantic symbols, channel allocation, and user power.

The work in [75] addressed the concept of compression ratio in the context of SC, particularly in the transmission of image and text semantic information. The compression ratio (C) was a crucial factor in the S-Rate of image semantic transmission. It represented the average number of semantic symbols per image ι. For a fixed compression ratio, CL_ι (where L_ι was the average pixel number in a single image) represented the average number of semantic symbols per image. This was crucial in the context of the paper's proposed NOMASC scheme, which aimed to enhance the performance of SC. The paper also introduced a new metric, the semantic metric, which was used in the definition of the S-Rate for image semantic transmission. The semantic rate of image semantic transmission (S-Rate) was defined as

$$\Gamma_\iota = \frac{WI_\iota}{CL_\iota} \cdot \xi_C^\iota(\gamma)\,(\text{suts/s}) \tag{8}$$

where I_i was the average amount of semantic information carried in a single image. This definition coupled the semantic rate with the transmission accuracy, representing the average amount of successfully transmitted semantic information per second.

In NOMASC system, the 'ergodic semantic rate' is a key metric [80]. This rate captures the average performance of SC across all channel conditions. Unlike a single snapshot of performance, it offers a comprehensive view of the system's capabilities. The ergodic semantic rate is derived from the instantaneous semantic rate $S(v)$, influenced by time allocation,

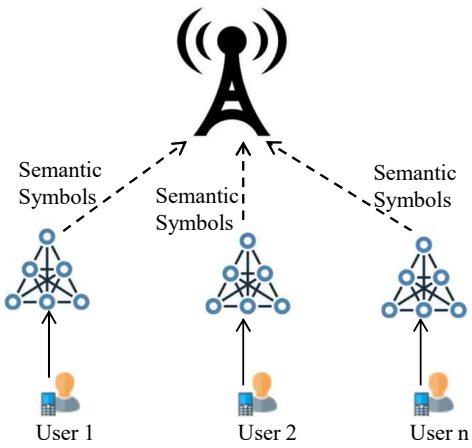


Fig. 4: Scenario Diagram of Semantic-aware Networks [8]

communication method (semantic or traditional), and channel conditions. The ergodic rate is the expected value of this instantaneous rate over all fading states:

$$\text{Ergodic Rate} = \mathbb{E}[S(v)] \tag{9}$$

Key factors affecting the ergodic semantic rate include transmit power, which can increase rate but also cause interference, and the choice of communication method. The study in [80] shows a flexible approach, switching between semantic and traditional methods to optimize performance.

2) QoS and QoE: In this section, we explore optimal resource management in SC systems, focusing on QoS and QoE. Unlike traditional systems that minimize technical errors, SC prioritizes the semantic value of information. This calls for new resource allocation frameworks that consider both QoS metrics like transmission rate and QoE factors like user satisfaction. Specifically, the QoS aims to optimize transmission rate, delay, and throughput, and the QoE focuses on user satisfaction, clarity, and fluency. In the following, we discuss the algorithms of resource allocation, user association, and power allocation for QoS and QoE in SC.

Xia *et al*. [77] addressed user association (UA) and bandwidth allocation (BA) challenges in intelligent SC within heterogeneous networks. They proposed a two-stage solution: the first stage used stochastic programming to achieve a deterministic objective with semantic confidence, and the second stage employed a heuristic algorithm to optimize UA and BA. This method aimed to maximize STM as described in Eq. (1), under KB matching and wireless bandwidth constraints. The results showed the proposed solution consistently outperformed baseline algorithms, even at high semantic confidence levels. Higher knowledge matching degrees between users and base stations further improved STM performance. Future work could enhance this solution and explore its application in various network scenarios.

Yan *et al*. [8] focused on optimizing resource allocation for text transmission in SC, aiming to maximize overall semantic spectral efficiency (S-SE). They highlighted the inadequacy of traditional spectral efficiency metrics for semantic information and proposed S-SE as described in Eq. (3), to measure the rate of successfully transmitted semantic information per unit bandwidth. To define S-SE, they introduced the semantic transmission rate as described in (2), referring to the effectively transmitted semantic information per second. As shown in Fig. 4, the scenario assumes multiple users sending semantic symbols to a base station. By formulating the resource allocation problem as an S-SE maximization, the authors employed exhaustive search methods and the Hungarian algorithm to find optimal solutions. Their findings demonstrated that SC systems could achieve higher S-SE than 4G and 5G systems, particularly when encoding words required more than 19 bits on average. Increasing the bits needed for encoding to over 27 bits with 10 dBm transmit power allowed SC systems to surpass even ideal traditional systems.

Building on these foundational works, Zhang *et al*. [76] focused on dynamic resource allocation for task-oriented SC systems using a Deep Deterministic Policy Gradient (DDPG) approach [81]. Their system involves intelligent devices performing feature extraction, semantic compression, and channel encoding for captured images, which are then uploaded to an edge server for processing and computation. The inference results are fed back to the devices for further action. This method jointly optimizes bandwidth, power, and semantic compression ratios, balancing task accuracy and execution frequency. Kang *et al*. [82] introduced a personalized saliency fused SC (PERSF-SC) framework for UAV image-sensing tasks, streamlining the communication process by transmitting only the most relevant data based on user interests. Additionally, Du *et al*. [83] proposed a SC framework for digital agriculture, optimizing power allocation to enhance the efficiency of transmitting semantic information in virtual apple orchards. In the context of video conferencing, Jiang *et al*. [84] introduced semantic video conferencing (SVC) to maintain high resolution under bandwidth constraints by transmitting key points representing motions. They proposed an incremental redundancy HARQ system to adapt to varying channel conditions and optimize bit consumption.

Further advancing resource management, studies by Farshbafan *et al*. [85] and Wang *et al*. [86] leveraged KG to optimize resource allocation. Farshbafan *et al*. proposed a hierarchical semantic KB structure and introduced "belief efficiency" to enhance communication resource efficiency, emphasizing the efficient utilization of "beliefs" for effective SC. Wang *et al*. employed a reinforcement learning algorithm with an attention network to optimize resource allocation and semantic information selection, significantly improving data transmission efficiency.

In the context of emerging wireless applications such as virtual reality, personalized healthcare, autonomous driving, and the Internet of Everything (IoE), satisfying a wide range of QoE requirements has become crucial, especially given the large volumes of data. There are some research works focused on QoE [73], [87]–[89]. Yan *et al*. [73] explored resource allocation in SC systems, specifically for text transmission. They aimed to maximize the overall semantic spectral efficiency (S-SE) of all users by addressing the challenge of quantifying the transmission rate of semantic information in

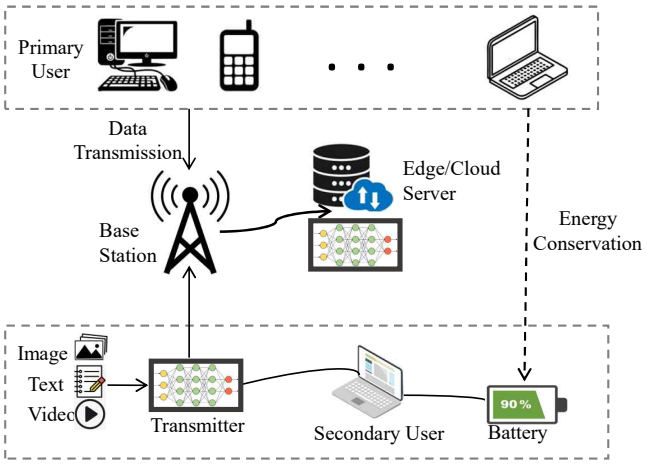


Fig. 5: Scenario Diagram of Paper [87]

communication networks. To tackle this, they proposed a novel QoE model as shown in Eq. (7). The QoE-aware resource allocation problem was formulated in terms of the number of transmitted semantic symbols, channel assignment, and power allocation, and was decoupled into two subproblems: optimizing the number of transmitted semantic symbols with given channel assignment and power allocation, and solving the channel assignment and power allocation subproblem using exhaustive search and a low-complexity matching algorithm. Their findings demonstrated the effectiveness and superiority of the proposed method in improving overall QoE, outperforming the random method and approaching the upper bound significantly.

Zhang *et al*. [87] addressed the problem of resource allocation in TOSCN for both uplink and downlink scenarios. Their system model, shown in Fig. 5, involved multiple primary users and one secondary user sharing spectrum resources using the NOMA protocol. They proposed a two-tier DRL framework integrating Deep Q Network and DDPG [81] to optimize the allocation of time slots, transmission power, and semantic compression ratio, aiming to maximize long-term QoE. The results showed that their method effectively allocated resources, leading to improved performance in a simulated TOSCN network, with QoE as a key performance metric.

Further advancing QoE optimization, Wang *et al*. [88] focused on maintaining high QoE in dynamic network conditions by adapting resource and power allocation in real-time. They designed an online algorithm, the soft actor-critic communication transformer (SACCT), to maximize total follower QoE gain while minimizing provider energy consumption. The algorithm made real-time decisions on encoding and transcoding bitrates, uploading power, and transcoding frequency based on current observations. Their results showed that the SACCT framework effectively maintained high QoE in live streaming scenarios by adapting to varying wireless channel conditions and bitrate requests. Similarly, Du *et al*. [89] aimed to deploy metaverse [90] ultra-reliable and low-latency communications (URLLC) services in wireless multiple-input and multiple-output networks. The problem addressed was the difficulty in providing a personalized immersive experience, a distinctive feature of the metaverse, using conventional URLLC. The authors proposed a novel metric, Meta-Immersion, for defining QoE in the metaverse, taking into account both objective network performance and subjective user attention values. Using an attention-aware rendering capacity allocation algorithm, they demonstrated that their URLLC attention-aware allocation scheme could increase Meta-Immersion by an average of 20.1%. compared to conventional URLLC schemes.

3) Energy Efficiency: The SC system must be developed to achieve energy efficiency under resource constrained condition. The efficient management of network resources is crucial for this purpose, as it not only extends the lifespan of the system for long-term operation but also enhances its reliability. Furthermore, efficient energy use has broader implications, including the reduction of carbon emissions and the mitigation of the impact on climate change. Several works have focused on energy efficiency in SC, including those by [10], [70], [74], [82], [91].

The intricacies of UAV image-sensing-driven task-oriented SC scenarios were explored by the authors in [82]. This research took a distinctive approach by emphasizing energy efficiency. They introduced an energy-efficient task-oriented SC framework employing a triplet-based scene graph for image information, ensuring that only the most pertinent information was transmitted, thereby conserving energy. A significant highlight was the introduction of a personalized semantic encoder tailored to user interests. This encoder ensured that the transmitted data was in perfect alignment with the user's requirements, avoiding unnecessary data transmissions and reducing communication overhead. In a traditional setup, a UAV would transmit 59 images to three subscribers, totaling 224.8MB of data transfer. However, with the PERSF-SC approach, the UAV selectively sent only 64 images to specified subscribers, significantly reducing bandwidth use and energy consumption during transmission [82].

Resource management is crucial for efficient energy use in SC systems. The authors in [74] conducted research on resource allocation and semantic information extraction in wireless network environments. They modeled scenarios where semantic information was extracted from large-scale data at base stations and transmitted as small-sized semantic information to users. Each user reconstructed the original data based on common knowledge built through collaborative learning. Probability graphs were employed at base stations to extract multi-level semantic information. In downlink transmission, a speed-splitting approach was adopted. Due to limited wireless resources, both computational and transmission energy were considered. An alternating algorithm was proposed to solve this optimization problem. Through simulation experiments, it was confirmed that the rate splitting multiple access-based approach proposed in [74] outperformed frequency division multiple access and NOMA-based approaches in terms of performance.

Several works have focused on optimizing MEC-enabled SC systems and adaptive compression techniques. The authors in [70] proposed an end-to-end SC architecture enabling users

to extract, compress, and transmit meaning. They introduced the adaptable semantic compression approach for compressing semantic information based on task importance given network conditions. The solutions for optimal compression ratio, resource allocation, and user selection were developed, achieving substantial data size reductions and improved task success rates. Additionally, the work in [10] proposed an SC system based on DNN models, introducing an MEC-based partitioned learning approach to reduce data traffic and energy consumption while increasing the number of supported wireless devices. The challenges of executing computationally intensive applications locally were addressed by the authors in [91], who enabled users to offload computation tasks to edge servers. This approach reduced computational energy consumption and enabled fast task execution through an approximation policy optimization-based multi-agent reinforcement learning algorithm.

4) Spectral Efficiency: The advent of 6G networks promises to revolutionize communication systems not just with faster speeds but with a fundamental shift towards ubiquitous AI and edge intelligence. Edge intelligence, derived from MEC, brings storage and processing closer to the data source, typically on small base stations, rather than distant clouds. This proximity is expected to significantly empower AI-driven MEC in 6G networks, offering massive processing capabilities and efficient data gathering [92], [93]. However, the increased intelligence and autonomy of devices will generate staggering volumes of data, necessitating high-performance connectivity that ensures low latency and manages anticipated network congestion and spectrum scarcity challenges [41]. SC has emerged as an effective solution for this issue. In contrast to Shannon-based communication, SC redefines spectral efficiency as semantic spectral efficiency. As a result, algorithms have been proposed to allocate resources by considering the transmission volume of semantic information [8].

As we delve deeper into SC, we encounter the dual challenge of efficient semantic information extraction and managing multi-user transmissions. With 6G expected to support connectivity densities up to 10^8 devices per km^2, enhancing spectral efficiency becomes essential. Effective spectrum management is crucial for regulating spectrum sharing and mitigating interference. The work in [94] tackles these challenges by presenting a CNN-based encoder-decoder architecture designed to enhance spectral efficiency in interference-heavy environments. This framework ensures accurate transmission and reconstruction of semantic information and employs an astute matching game with externalities to refine channel allocation, thereby amplifying the mean semantic rate per user.

Maximizing spectral efficiency is further facilitated by NOMA technology, which enables multiple users to share the same frequency band for simultaneous communication [95]. Integrating NOMA technology and semantic-aware resource allocation frameworks, as highlighted in studies such as [75], [96]–[98], is vital for the success of SC in 6G environments. These studies advocate for a shift from traditional bit-based communication systems to methodologies that ensure more efficient in multi-user. The research in [75] builds on this by proposing a NOMASC system for non-orthogonal semantic transmission across various datasets and data modalities. Studies such as [96], [97] explore the application of NOMA technology to SC, introducing a heterogeneous semantic and bit multi-user framework for efficient multi-user communication in various scenarios. A significant highlight of these studies is the introduction of the semi-NOMA concept, a unified multiple access scheme designed to facilitate heterogeneous semantic and bit multi-user communication. The authors emphasize the potential of NOMA when integrated with SC to support this heterogeneous framework. Fig. 6 provides a detailed insight into the Semi-NOMA approach, specifically designed for heterogeneous semantic and bit communications. The non-orthogonal sub-band operates on the NOMA principle, accommodating multiple simultaneous transmissions at varied power levels. Conversely, the orthogonal sub-band is reserved for specific users, adhering to the orthogonal multiple access (OMA) approach. This Semi-NOMA strategy harmoniously blends NOMA's flexibility with OMA's interference-free communication. Additionally, the diagram underscores the significance of both bit rate and semantic rate. While the bit rate denotes data volume transmitted over time, the semantic rate gauges the conveyed semantic information's richness. This comprehensive approach demonstrates the potential of Semi-NOMA in addressing the multifaceted communication demands of modern systems, from rich semantic content to traditional data bits.

Modern SC applications, such as XR devices, demand advanced metrics and resource allocation strategies that traditional communication metrics cannot meet. The study in [98] introduces the Age of Incorrect Information (AoII) metric, which captures both error-based and Age of Information (AoI)-based performance features. This metric provides a comprehensive view of system performance, particularly in NOMA-aided XR devices, ensuring efficient resource allocation and improved performance in task-oriented SC. XR devices offer immersive digital environments but present challenges like ultra-massive access and real-time synchronization, which traditional metrics can't address. To bridge this gap, the AoII metric was developed, capturing both error and AoI features to provide a holistic view of system performance. The study constructed a multi-user uplink NOMA system, analyzed its performance using AoII, and derived average semantic similarity and closed-form expressions for packet AoI. A non-convex optimization problem was formulated, considering semantic rate, transmit power, and status update rate, and solved with a linear search-based algorithm. Simulation results demonstrated the effectiveness of the AoII metric in evaluating transmission performance.

A summary of existing works of semantic-aware resource allocation in Section III-C is presented in Table VI.

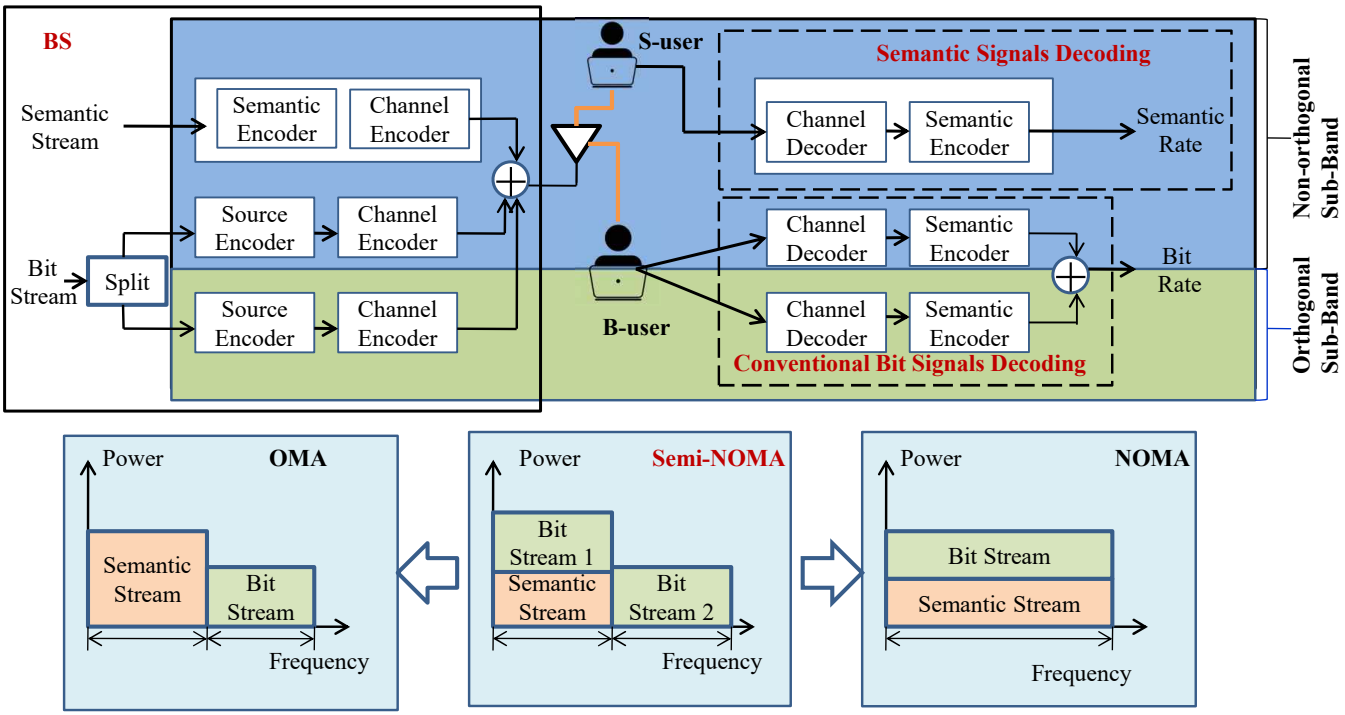


Fig. 6: Semi-NOMA for Heterogeneous Semantic and Bit Communications [97]

TABLE VI
SUMMARY OF SEMANTIC-AWARE RESOURCE ALLOCATION

Ref.	Algorithm	Direction	Considered Objectives				Data Type
			QoS	QoE	EE	SE	
[77]	RA-UA	DL	O	X	X	X	Text
[86]	RA	DL	O	X	X	X	Text
[98]	RA-PA	UL	O	X	O	X	Text
[8]	RA	-	O	X	X	O	Text
[96]	RA-PA	DL	O	X	X	O	Text
[83]	RA-PA	DL	O	X	X	X	Image
[76]	RA-PA	UL-DL	O	X	X	X	Image
[84]	RA-PA	DL	O	X	X	X	Video
[82]	RA-PA	DL	O	X	O	X	Multimodal
[85]	RA	-	O	X	X	X	Multimodal
[73]	RA-PA	UL	O	O	X	X	Multimodal
[87]	RA-PA	UL	O	O	X	O	Image
[75]	RA-PA	DL	X	O	O	X	Multimodal
[99]	RA-PA	UL	X	O	O	X	Multimodal
[88]	RA-PA	UL	X	O	O	X	Video
[89]	RA-PA	UL-DL	X	O	O	X	Image
[94]	RA	-	X	X	O	O	Image
[91]	RA-PA-UA	UL	X	X	O	X	Text

Notation: In the table, "RA" means the resource allocation, "PA" means the power allocation, "UA" means the user association, "UL" means the uplink and "DL" means downlink.

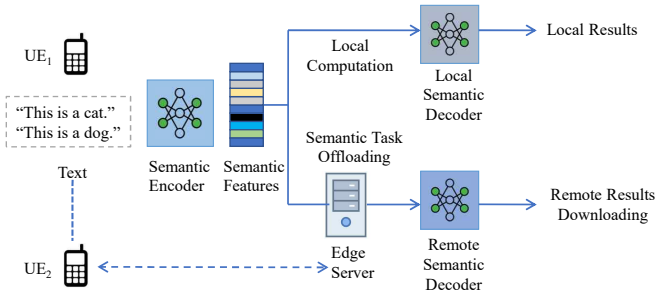


Fig. 7: System Model for Semantic Task Offloading [91]

D. Semantic-aware MEC Orchestration

1) Semantic-aware Offloading: The rapid evolution of communication networks has ushered in an era where the sheer volume of data, often termed as big data, necessitates a more integrated approach to handling and processing. With the rise of MEC, there is a pressing need to efficiently manage and optimize resources to cater to the demands of modern applications. Modern communication networks are envisioned as distributed systems that not only transfer data quickly but also process, store, and retrieve it efficiently. This is particularly crucial at the edge, where resource constraints are more pronounced compared to centralized cloud systems. MEC ensures that data are processed and available with minimal delay, meeting the demands of low-latency applications such as IoT, automated driving, and augmented reality (AR) [100]. In such MEC-based semantic-aware networks, issues like task offloading and caching have become prominent. In MEC-based SC, semantic encoders and decoders are used to offload semantic tasks to edge servers. These systems face a fundamental trade-off in deciding whether tasks should be processed locally on user devices or offloaded to external servers equipped with semantic decoders and applications. When semantic tasks are processed locally on user devices, the primary drawback is the substantial energy consumption required to handle complex computations, which can deplete device batteries quickly and reduce operational efficiency. On the other hand, offloading tasks to remote servers reduces local energy usage but increases the consumption of network resources. Thus, the decision between local processing and remote offloading must carefully consider the balance between energy efficiency and network resource utilization, aiming to optimize overall system performance and user experience.

Ji *et al.* [91] addressed these challenges by introducing a semantic-aware task offloading system with a focus on enhancing the energy efficiency of user equipment in edge networks. The system extracted the intrinsic semantic information of tasks and strategically offloads them to edge servers. This innovative approach was tailored to reduce the computational energy consumption of user devices, ensuring a more sustainable and efficient operation. Fig. 7 illustrated the semantic task offloading mechanism within a network. Two user equipments (UEs), UE_1 and UE_2, processed distinct text inputs: "This is a cat" and "That is a dog", respectively. These inputs were transformed by a semantic encoder into compact semantic features. The UEs then decided whether to process these features locally or offload them to an edge server for remote processing. Local processing reproduced the original text, while offloading to the edge server resulted in the remote results being returned to the UEs. Extending their work, Ji *et al.* [101] explore multimodal scenarios to further enhance the implementation of SC for task offloading. This involves optimizing the QoE while considering user preferences, making it the first study to perform joint optimization of computational and communication resources in semantic-aware networks with multimodal tasks.

2) Semantic-aware Caching: Caching is another critical aspect of SC systems in MEC environments, especially as the demands of 6G applications grow. The study in [102] explores the coordination of cache and computing resources within a multimodal SC-assisted MEC system. By introducing a cache-enhanced offloading scheme, the study aims to minimize the overall cost of computation of the system. The authors formulated a bidirectional caching task model and developed a content popularity-based DQN caching algorithm to make near-optimal caching decisions. This approach effectively reduces system cost and improves cache hit rate, laying a foundation for future research in cache-computation collaboration schemes [102]. Furthermore, [103] proposes a KB deployment mechanism based on edge caching to support SC applications. This architecture integrates intelligent devices with KBs located at both the cloud and the edge, ensuring efficient semantic operations. Given the constraints of edge servers, a subset of the semantic KB is stored, requiring an efficient deployment strategy. Initially, a semantic KG is formed using historical device request data and stored in the cloud. When devices send semantic understanding requests, the edge responds by executing a subgraph query. Over time, the cloud KB updates, and the edge's KB can also be refreshed. This approach underscores the importance of cache-enabled KBs in optimizing SC for future 6G networks.

3) Semantic-aware Control: The rapid evolution of communication networks and the onset of 6G technology highlight the critical need for semantic control and filtering in communication systems. The paper [26] emphasizes the importance of integrating information semantics, not just as the meaning of messages, but as their significance in relation to the purpose of data exchange. The authors highlight the necessity of moving away from the traditional model where humans selected the data and the network ensured its accurate and timely delivery, driven by the need for automated decisions within a sense-compute-actuate cycle. Instead, they advocate for a design that delivers the right and significant piece of information to the appropriate computation or actuation point at the right time, which is essential for the scalability of these systems. The paper [26] delves into the theoretical foundations of redesigning the entire process of information generation, transmission, and usage for networked systems, emphasizing the development of advanced semantic metrics for communication and control systems, optimal sampling theory, and related concepts. Additionally, the concept of a semantic-effectiveness plane, discussed in [104], represents a significant evolution in communication architectures by

offering standardized interfaces for information filtering and control across all layers of the protocol. This approach facilitates efficient data transmission and semantic-aware control, underscoring the need for semantic communication systems that are responsive and adaptable to the demands of modern networks.

4) Semantic-aware C4 Orchestration: The integration of communication, computation, caching, and control (C4) resources in SC systems is vital for modern communication networks. The study in [105] highlights the necessity of a C4 framework within MEC to optimize bandwidth consumption and network latency, demonstrating effective reductions in bandwidth use and latency. The SCORING project, proposed in [106], emphasizes a collaborative computing, caching, and networking paradigm facilitated by SDN/NFV layers, enabling resource sharing across clouds, core networks, edge servers, and end devices. This approach incorporates network slicing and MEC-enabled microservices to meet stringent performance requirements through AI and machine learning integration. SCORING's architecture, featuring a Management and Orchestration plane, is designed to manage MEC-enabled microservices and integrate computing, storage, and networking, ensuring networks are adaptable and ready for future demands. The development of semantic KBs is also crucial for supporting SC in resource-constrained environments, underscoring the need for a robust C4 framework in modern communication networks.

E. Lessons Learned and Summary

The research covered in Section III-B highlights the evolving landscape of semantic-awae information transmission system. In Section III-B1, the emphasis is placed on designing semantic-aware transceiver systems to enhance task-oriented and system-oriented performance or fidelity. Key approaches involve defining a semantic loss or distortion function and minimizing it to make semantic encoders and decoders that effectively extract semantic information. Additionally, the trade-off between the semantic encoder's rate and semantic distortion was analyzed. This section explored how to design SC systems for various types of data such as speech, visual, and text, as well as for integrated multimodal data. Furthermore, research on semantic noise was conducted to define and design robust semantic encoders and decoders. Advanced techniques like KG and Bayesian networks were used to define new semantic concepts through knowledge representation units. Studies also investigated designing semantic transceivers with generalizable intelligence using approaches like domain adaptation. Section III-B2 examines methods for dynamically controlling the rate of the semantic encoder by considering the trade-off between the rate and semantic distortion based on given network conditions. For example, adaptive methods using multi-bit length selection and progressive semantic HARQ schemes were proposed to reduce communication costs and semantic errors. Section III-B3 explores research that integrates adaptive control and resource allocation to optimize performance across various network scenarios. By jointly controlling the amount of transmitted semantic information and allocating resources efficiently, this approach aims to enhance the overall performance of SC system.

In Section III-C, we explore various aspects of resource management and performance optimization in SC systems. We learned that traditional Shannon-based communication metrics need to be redefined from a semantic perspective, focusing on metrics like semantic rate, semantic transmission rate, and QoE to optimize resource allocation and performance. Key approaches include stochastic programming, heuristic algorithms, and DRL for dynamic resource management, with the aim of improving both QoS and QoE. Additionally, energy efficiency is critical, with strategies focusing on transmitting only the most relevant information and optimizing resource use to extend the lifespan of the system and reduce the environmental impact. The importance of managing spectral efficiency in 6G networks was highlighted, with innovations like NOMASC systems and the semi-NOMA concept balancing flexibility and interference-free communication. We also examined how traditional multi-access schemes need modifications from a SC perspective, and how resource management can be handled heterogeneous situations with bit and semantic communication. These insights collectively underscore the need for efficient, semantically aware resource management to meet the complex demands of modern communication systems.

In Section III-D, we learned the necessity of integrating C4 resources to optimize SC systems. The rise of MEC emphasizes the need for efficient management of these resources. Key insights include the importance of semantic-aware caching, offloading, and sampling control techniques. For instance, semantic task offloading can balance energy efficiency and network resource utilization, while caching enhances system performance by storing frequently accessed data closer to the edge. Moreover, joint C4 orchestration is essential to manage the complexities of SC environments, ensuring low latency and high efficiency in data processing and transmission. Studies highlighted the use of advanced frameworks and algorithms, such as DRL and knowledge-based caching, to achieve these goals and support the stringent requirements of modern applications such as IoT and AR.

IV. Security and Privacy in SC

In Section IV, we analyze potential attacks targeting key components of SC systems, including the encoder/decoder, KB, and transmission channel. By identifying vulnerabilities in each part, we aim to propose strategies to enhance the security and privacy of SC systems. SC systems are susceptible to various attacks due to the complex interplay of their components. Below is a detailed description of attacks categorized by the affected component, along with the sections of the paper that discuss these attacks.

The encoder/decoder component faces multiple security threats, including poisoning attacks, which introduce corrupted data into the training set to degrade the performance of the encoder/decoder. Backdoor attacks involve inserting malicious triggers into the encoder/decoder to manipulate the output when specific conditions are met. Adversarial example attacks

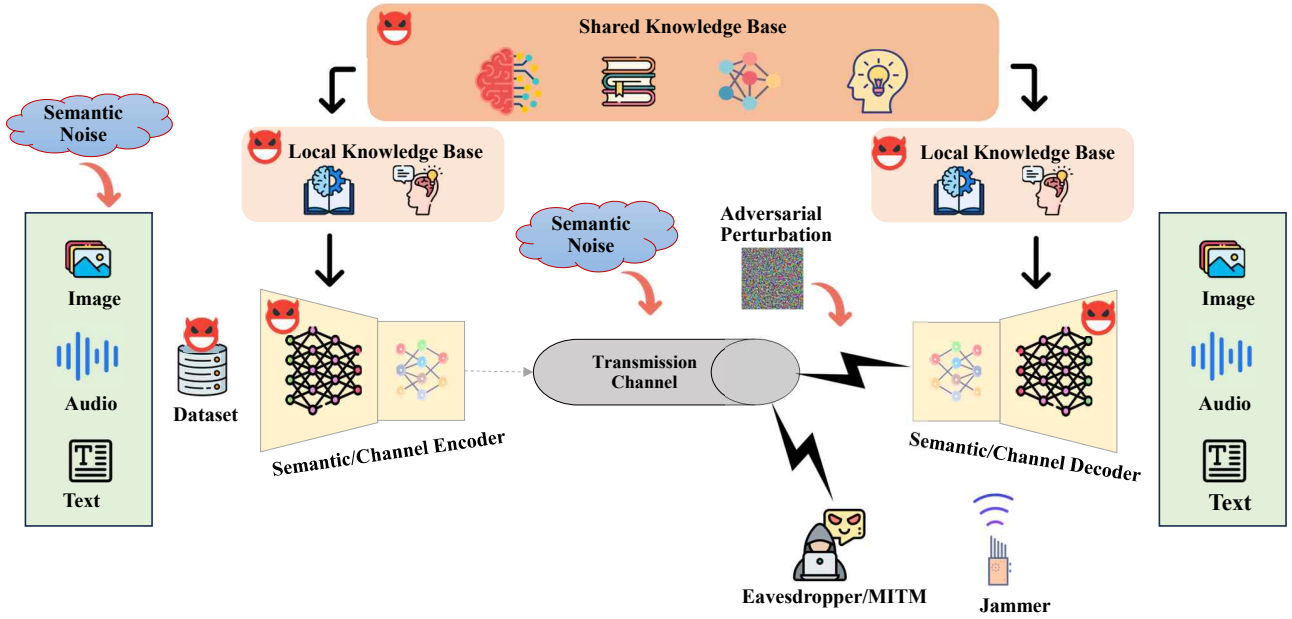


Fig. 8: Overview of Attacks in SC Systems

use slightly modified inputs to deceive the encoder/decoder into producing incorrect outputs. Model inversion (MI) attacks aim to extract sensitive information from the model, while membership inference attacks attempt to determine if a particular data point was part of the training set. Additionally, semantic noise introduces misleading or incorrect semantic information, further compromising the system's performance. These attacks are discussed in detail in Section IV-A of the paper.

The KB, a critical component for storing and managing semantic information, is also vulnerable to attacks. Data poisoning involves attackers injecting malicious data into the KB, leading to incorrect semantic interpretations. Data tampering attacks occur when unauthorized users gain access to the KB and alter the stored information, which can result in the dissemination of false or misleading data. These attacks are discussed in detail in Section IV-C of the paper.

The transmission channel is susceptible to several types of attacks. Eavesdropping involves unauthorized parties intercepting communication to gain access to sensitive information. Jamming attacks disrupt communication by overwhelming the transmission channel with noise or false signals. Man-in-the-middle attacks occur when attackers intercept and potentially alter the communication between the sender and receiver. Physical layer adversarial attacks exploit weaknesses at the physical transmission level to introduce errors and distortions, compromising the integrity and reliability of the communication. These attacks are detailed in Section IV-C of the paper.

Fig. 8 provides a comprehensive overview of attacks in SC system, highlighting the vulnerability of components such as the encoder/decoder, KB, and transmission channel.

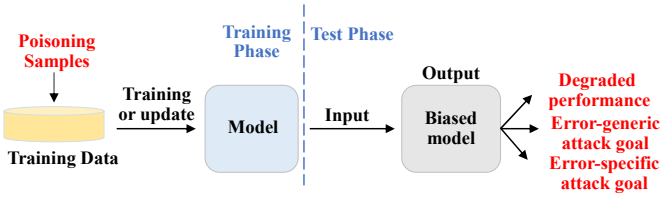


Fig. 9: Overview of Poisoning Attacks [107]

A. *Encoder/Decoder Security*

In traditional communication systems, the encoder and decoder operate based on predefined rules and algorithms, facilitating direct mapping and transmission of raw data with minimal processing without involving DL or ML in the encoding or decoding processes. On the other hand, SC systems utilize ML to learn underlying patterns and relationships between raw data and its semantic representation. However, the susceptibility of ML models to various types of attacks [107], particularly in high-stakes applications like medical diagnosis and autonomous vehicles, raises significant concerns. Even minor errors in a model's decision-making process can lead to disastrous consequences, highlighting the urgent need to ensure the security of the encoder/decoder in SC systems. Numerous types of attacks have been identified that can compromise the security of ML models, including those utilized in SC systems [16].

1) Poisoning Attacks: As depicted in Fig. 9, a poisoning attack involves an attacker deliberately manipulating the training data used for the encoder/decoder to degrade its performance, either by causing incorrect predictions or introducing bias [107], [108]. This can be done by adding malicious corrupted data to the training set in various ways, such as modifying existing data or adding new data designed to influence the

model's decisions [108]. In the context of wireless end-to-end image transmission systems, SC has emerged as a promising approach to conserve bandwidth [109]. A study in [110] proposed a technique for wireless image transmission using DeepJSCC, emphasizing the importance of well-trained datasets at both the transmitter and receiver for reliable and efficient performance. However, these training datasets are not immune to security vulnerabilities. One notable vulnerability is poisoning attacks, as demonstrated by Chen *et al*. [111]. They proposed a type of attack called DeepPoison, which involves the insertion of specific triggers into benign training data, causing the network to misclassify certain inputs. To generate these poisoned training samples, the proposed method utilizes a generator and two discriminators, ensuring that the poisoned samples are indistinguishable from their benign counterparts.

Xie *et al*. also investigated the susceptibility of DNNs to poisoning attacks, particularly focusing on video recognition models [112]. Their study involved manipulating the training data using a trigger pattern to induce misclassification of data instances. To enhance stealthiness and minimize visual changes, the paper introduced a novel 3D poisoning attack framework. This framework leveraged a computer graphic primitive to construct the poisoning trigger, achieving significantly reduced visual alterations in the manipulated videos.

In speech transmission systems, SC was also leveraged to reduce data consumption. Weng *et al*. introduced DeepSC-ST, a DeepSC system specifically designed for speech transmission [113]. In this system, the input spectrum was transformed into text-related semantic features at the transmitter. These features were then extracted and transmitted using a joint semantic-channel encoder, while the receiver utilized the received semantic features to recover the transmitted text. Additionally, speech synthesis was performed at the receiver by incorporating the recognized text and speaker information into a semantic-channel encoder, enabling the reconstruction of speech signals. The proposed model exhibited potential for the development of digital voice assistant systems in various contexts, such as home, car, or smartphone applications, offering heightened convenience in our daily lives. However, it was crucial to address security concerns associated with the training process. Automatic speech recognition (ASR) systems necessitated vast amounts of training data, often collected from potentially untrustworthy sources. This posed a challenge as malicious actors could introduce poisoned data, thereby compromising the integrity and performance of the ASR system. Vigilance was essential to mitigate these risks and ensure the security of the speech transmission system.

For instance, in [114], VENOMAVE was introduced as the first training-time poisoning attack designed specifically for speech recognition systems. This attack aimed to manipulate the system's training data to generate an incorrect transcription of a targeted audio waveform, as determined by the attacker. The objective was to deceive the system into recognizing specific commands, such as "open the door," even when the user was saying something entirely different. Unlike traditional approaches that manipulated input utterances, VENOMAVE achieved its desired outcome by tampering with the system's training data. The results of the poisoning attack demonstrated that a small number of poisoned samples could effectively compromise the system with a high success rate.

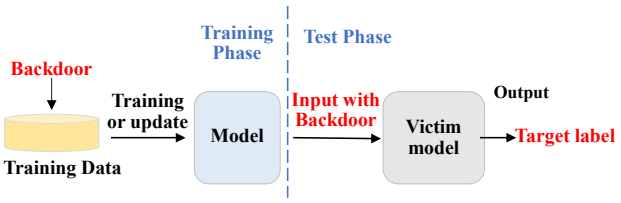


Fig. 10: Overview of Backdoor Attacks [107]

Similarly, in healthcare applications, poisoning attacks are of great concern, particularly due to the limitations associated with directly manipulating training data. In [115], the significance of such attacks was emphasized, as they could lead to serious misclassification errors in medical datasets, especially when classifying cancer and disease samples.

2) Backdoor Attacks: As depicted in Fig. 10, backdoor attacks pose a significant threat to SC systems, occurring during both the training and testing phases of the encoder/decoder. In the training phase, malicious attackers can exploit vulnerabilities by introducing triggers into the training data. These triggers manipulate the learning process, causing the encoder or decoder to misinterpret the meaning of the data during subsequent communication. During the testing phase, the attacker can inject additional poisoned samples into the input data, causing the encoder or decoder to make erroneous classifications and produce misleading semantic representations. These attacks involve an adversary embedding triggers into a limited number of training samples and altering the corresponding labels to a predefined target [107]. For instance, during the training phase, the adversary introduces triggers, such as a "plus sign", into specific positions of input images and modifies their labels to match the target label. Subsequently, in the testing phase, the adversary activates these triggers by providing poisoned samples as input to the encoder or decoder of the SC system. As a result, the backdoor attack effectively manipulates the semantic information conveyed by the poisoned input samples to achieve a desired target meaning [116].

Backdoors can occur in almost every stage of the ML pipeline. Zhang *et al*. proposed a new backdoor attack method called Poison Ink that utilized the image structure as the carrier of poison information to generate trigger patterns and leveraged a deep injection network to hide the trigger patterns in the cover images invisibly [117]. Jia *et al*. discussed the vulnerability of pre-trained image encoders in self-supervised learning to backdoor attacks [118]. The pre-trained image encoder could then be used as a feature extractor to build downstream classifiers for many downstream tasks with a small amount of or no labeled training data. The paper proposed BadEncoder, which injected backdoors into a pre-trained image encoder, allowing the attacker to make attacker-chosen predictions for inputs embedded with an attacker-chosen trigger.

A recent study introduced the visible, semantic, sample-specific, and compatible (VSSC) trigger to address limitations

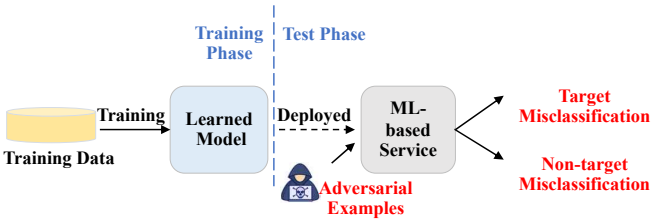


Fig. 11: Overview of Adversarial Example Attacks [122]

in DNN backdoor attacks [119]. This approach used a stable diffusion model to generate corresponding semantic objects that seamlessly integrated with the original images, creating realistic poisoned images. The VSSC trigger was found to be universally effective, stealthy, and robust, regardless of the poisoning ratio.

In addition to image classification, backdoor attacks also pose a significant threat to text classification tasks, such as in LSTM-based systems [120]. The paper proposed a three-phase approach: generating poisoning samples, training the model with poisoned data, and activating the backdoor. This attack was conducted in a black-box setting, with the adversary having limited knowledge of the model structure and training algorithms, except for a small amount of training data. Experimental results showed a 96% success rate with a 1% poisoning rate. In [121], a backdoor attack named BITE was introduced, embedding a backdoor into a victim model by providing poisoned training data. BITE established strong correlations between the target label and trigger words by iteratively identifying and injecting them into target-label instances through natural word-level perturbations. This instructed the victim model to predict the target label for inputs containing trigger words, forming the backdoor.

3) Adversarial Example Attacks: As depicted in Fig. 11, adversarial example attacks primarily target the testing or inference phase of the encoder/decoder. During this phase, the attacker manipulates the input data to appear innocuous to humans, but this leads to misclassification or incorrect outputs from the model. The goal is to exploit vulnerabilities in the model's decision-making process and cause it to make mistakes when confronted with these modified inputs [107]. Adversarial example attacks could potentially have devastating consequences for healthcare systems, particularly those reliant on ML systems.

A comprehensive study by Rahman *et al*. examined the security risks associated with ML systems for COVID-19 in medical IoT devices. Their findings revealed that models without defensive mechanisms against adversarial perturbations were susceptible to attacks [123]. Similarly, the impact of such attacks on autonomous driving systems is significant. Autonomous vehicles rely on accurate environmental perception for decision-making, and errors can have catastrophic real-world consequences. ML-based algorithms have shown potential in semantic segmentation, which can enhance SC by reducing data transmission requirements. For instance, only pixels corresponding to a car need to be transmitted instead of the entire image. However, adversarial examples can deceive the model by subtly altering car images to misclassify them as pedestrians, leading to dangerous actions such as sudden brakes or unsafe lane changes [124], [125].

Various techniques can be employed to generate adversarial examples in the domain of autonomous driving. Among these techniques, one commonly employed method is gradient descent. Gradient descent is an optimization algorithm utilized to identify the minimum of a given function. In the realm of adversarial examples, gradient descent is leveraged to determine the smallest perturbation that can induce an incorrect prediction by the model [126]. To minimize perturbation costs and enhance the effectiveness of adversarial attacks, Yang *et al*. introduced an approach called targeted attention attack. This method utilizes the concept of the Recurrent Attention Network to identify the most critical pixels in the input image. By optimizing a universal perturbation, the targeted attention attack technique aims to generate subtle perturbations that can be easily overlooked by human drivers but significantly increase the fooling rate when applied to a diverse set of test images [127].

The emergence of adversarial attacks in the field of image classification has sparked growing interest in investigating adversarial audio attacks on ASR systems. However, due to the inherent complexity of ASR structures, constructing precise audio adversarial examples that align perfectly with the desired target text transcription presents a significant research challenge. Generally, adversarial attacks on ASR systems can be categorized into three main types based on the attacker's knowledge: 1) white-box attacks, where the attacker possesses complete knowledge of the targeted ASR system, 2) gray-box attacks, where the attacker has limited knowledge about the system, and 3) black-box attacks, where the attacker has no prior knowledge of the system. Although existing research on adversarial attacks in ASR has focused on white-box attacks, it is important to note that this assumption does not always hold in real-world scenarios. In practice, attackers often have restricted access to only the output of the ASR system, lacking detailed insights into the underlying architecture or parameters [128].

In order to launch successful black-box attacks on high-dimensional input targeting models, attackers had to introduce imperceptible perturbations to the original examples, thereby manipulating the model to produce the desired target text. Wang *et al*. proposed a novel black-box attack technique called the Monte Carlo gradient sign attack, which significantly reduced the number of queries required to generate adversarial audio samples against ASR systems. The proposed method utilized a Monte Carlo tree to identify elements within the original audio sample that exhibited dominant gradients. Using a sampling gradient sign strategy and an iterative momentum strategy, the original sample was updated to create an adversarial example. The authors attributed the high query efficiency of their approach to the effective exploitation of the dominant gradient phenomenon. When subjected to the Monte Carlo gradient sign attack, the ASR system experienced incorrect speech transcription. The generated adversarial examples successfully deceived the ASR system, causing it to incorrectly recognize the altered speech content [129].

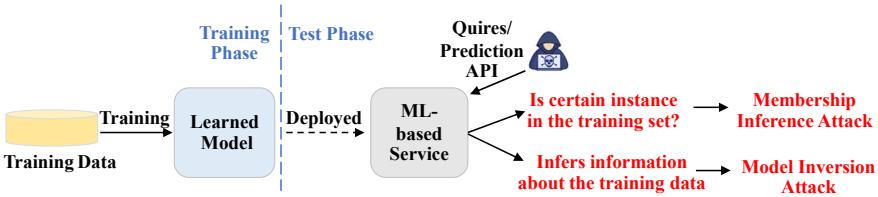


Fig. 12: Overview of Model Inversion and Membership Inference Attacks [107], [122]

4) Model Inversion Attacks: As depicted in Fig. 12, MI attacks involve querying the model with crafted inputs and analyzing the responses to infer details about the training data, which can include personal conversations, medical records, and financial data. MI attacks significantly impact SC systems by extracting sensitive or private information from ML models through careful observation of their outputs [107]. SC systems are particularly vulnerable to these attacks due to their reliance on identifying and understanding semantic patterns to generate meaningful responses. Successful MI attacks can expose private information and provide insight into the model's decision-making process, allowing adversaries to manipulate its behavior. This can lead to biased, misleading, or inappropriate responses, compromising the integrity and reliability of the communication.

The first proposed MI attack was introduced in the context of genomic privacy [130]. Fredrikson *et al*. addressed the privacy concerns related to ML models utilized in the field of pharmacogenetics, particularly in the context of personalized warfarin dosing. Pharmacogenetics involved using ML models to inform medical treatments based on a patient's genetic makeup and background. In their research, the authors demonstrated the feasibility of performing MI attacks to predict a patient's genetic markers given access to the ML model and some demographic information. This highlighted the potential privacy risks associated with using ML models in pharmacogenetics, as attackers could exploit these models to infer sensitive genetic information about individuals. Zhang *et al*. proposed an attack method called the generative MI attack [131], which could invert DNNs with high success rates. They used partial public information to learn a distributional prior via generative adversarial networks (GANs) and used it to guide the inversion process to reconstruct private image data from a state-of-the-art face recognition classifier. Their MI attacks were effective even when public datasets did not include the identities that the adversary aimed to recover.

Typically, an MI attack required attackers to query the auxiliary datasets entirely to gather information about the target inference models. However, this approach could be inefficient and raise concerns related to transferring large datasets to online services, as well as potentially triggering active defense mechanisms by administrators. To address these challenges, an MI attack scheme proposed in [132] suggested an alternative approach. The scheme proposed utilizing latent information extracted from primitive models as high-dimensional features, thereby reducing the reliance on extensive querying of the auxiliary datasets. In addition to image data, MI attacks could also have an impact on textual data. The paper [133] introduced an MI attack specifically designed for text reconstruction in the context of text classification using transformers. The authors proposed an MI attack called Text Revealer, which was the first of its kind for text reconstruction in transformers-based text classification. The attack method aimed to reconstruct private texts that were part of the training data, using access to the target model. To achieve this, the authors utilized an external dataset and the GPT-2 language model to generate fluent text resembling the target domain. They then optimized the hidden state of the generated text using feedback from the target model, in order to perturb it optimally and reconstruct the private texts.

5) Membership Inference Attacks: As depicted in Fig. 12, the membership inference attack is a type of privacy attack that aims to determine whether a specific data point was used during a model's training phase [107]. The attacker, with access to the target model and its output predictions, exploits the model's behavior to infer membership information about the training dataset. By observing the model's predictions on multiple inputs, the attacker can analyze patterns to discern whether a particular data point was part of the training data [134].

SC systems often handle sensitive user data, and a successful membership inference attack can reveal whether a specific user's data was part of the training dataset, breaching their privacy. A study in [135] investigated the vulnerability of Clinical Language Models (CLMs) to membership attacks and estimated their privacy leakage. CLMs are trained on clinical data to improve performance in biomedical natural language processing tasks. The study assessed the risks of training-data leakage through white-box or black-box access to CLMs, employing membership inference attacks to estimate empirical privacy leaks for models like BERT and GPT-2. The results showed that membership inference attacks on CLMs could lead to privacy leakages of up to 7%, posing a risk to patient privacy.

Recommender systems are widely used in various services today, including online shopping and streaming platforms. Traditional recommender systems rely heavily on explicit user feedback, such as ratings or interactions, to generate recommendations. However, this approach faces challenges with the cold start problem, where new users or items have limited or no historical data. By incorporating semantic information, recommender systems can overcome this limitation and provide relevant recommendations even when explicit feedback is scarce or absent. However, recommender systems

often use sensitive user data, and potential data leakage can lead to severe privacy issues. The paper [136] proposed a novel method to quantify the privacy leakage of recommender systems through membership inference. Unlike traditional membership inference attacks on ML classifiers, this attack operates at the user level, where the adversary can only observe the ordered recommended items. To address these challenges, the authors proposed a method for representing users based on relevant items and established a shadow recommender to derive labeled training data for the attack model.

6) Semantic Noise: Semantic noise is a type of noise that causes misunderstandings of semantic information by introducing discrepancies between intended and received semantic symbols. The causes of semantic noise vary depending on the nature of the information source, such as text or image data. The fidelity of semantic information extracted and processed by the semantic transceiver is crucial for successful SC. However, semantic noise can disrupt this process, leading to disturbances in conveyed semantic information.

In textual information, semantic noise includes ambiguity, where slight modifications like synonym replacement, typos, or grammatical errors can mislead ML models. For instance, the sentence "The cat is sitting at the bank" can be ambiguous, with "bank" potentially referring to a financial institution or the land next to a river. This ambiguity, known as literal semantic noise, is classified as such because of its inherent uncertainty [137].

Regarding image data, [138], [139] discuss modeling semantic noise through adversarial samples. By introducing subtle perturbations to image data using the fast gradient sign method, these modifications can cause misinterpretations in SC systems' encoder and decoder, degrading model performance. This method leverages the gradients of the loss function with respect to the input image, directing small perturbations that increase the likelihood of misclassification. Additionally, semantic noise exists naturally. Capturing images of adversarial samples can lead to misclassification because models trained only on clean images lack robustness to noise. They focus on classifying pristine images and neglect the ability to handle noise, which is prevalent in real-world scenarios. Factors such as shadows, blur, and occlusion can affect images, making accurate predictions challenging [140].

B. Knowledge Base Security

SC relies on extracting the intended meaning of the transmitted information from the sender and accurately interpreting it at the receiver. This is achieved through the use of a matched KB, which serves as the foundation of SC and contributes to the subjectivity of semantic information. The unique background knowledge that individuals have influences how they perceive and describe the world, leading to variations in SC performance. To mitigate this, transceivers share their knowledge through a shared KB before transmitting data [14].

SC primarily focuses on extracting semantic features from multimodal signals and utilizing a KB to interpret these features in a form that is understandable to users. Certain source signals, such as video, audio, and haptic signals, inherently possess multiple meanings, making it challenging to recognize the intended meanings without relevant background knowledge or contextual information [141]. The transmitter and receiver each possess their own background knowledge, referred to as the local KB, which may contain different information. Consequently, the receiver's interpretation of transmitted semantic data may not always be entirely accurate. To address this issue, a common KB, also known as the shared KB, is established and shared between the transmitter and receiver. This shared KB aids the transmitter in extracting semantic information and enables the receiver to more effectively recover the underlying information from the received data. The establishment of a similar KB between the communication parties is crucial for effectively conveying the meaning of information in SC. Conditional mutual information between the KBs of the parties significantly enhances the performance of SC. Additionally, the dataset used to train the encoder and decoder is considered a form of shared background knowledge. It is vital for this dataset to accurately represent the type of information that will be transmitted within the SC system, ensuring that the encoder and decoder can faithfully capture the meaning of the transmitted information [142].

Moreover, the terms KG and KB are often used interchangeably. A KB serves as the building block for creating a KG, providing structured and unstructured data that can be extracted and represented as nodes and edges in the graph [143]. It is a semantic network that reveals the relationship among entities in the form of graphs, and it is constructed using top-down approaches, including information extraction, knowledge fusion, knowledge processing, and knowledge update. KG is used to transform triples into natural text, which is crucial to interpret the meaning of semantic information [144]. The KG finds applications in various domains, including question answering, recommendation systems, and natural language processing. However, the KB in SC system is not immune to potential attacks that seek to modify, steal, or perturb its data. These attacks can compromise the integrity and reliability of the information that is transmitted. Adversaries may attempt to manipulate the KB to alter the intended meaning of the communicated information, leading to misunderstandings or misinformation.

1) Knowledge/Data Poisoning Attacks: The integrity of the KB faces a looming threat in the form of poisoning attacks. These malevolent intrusions involve the deliberate insertion of corrupted data, with a particular focus on publicly accessible repositories. The quality of data within the KB significantly influences the extraction and recovery of semantic features. Therefore, the presence of tainted information seriously compromises the authenticity and availability of the entire dataset. Consequently, the credibility of the KB is at stake, potentially leading to significant disruptions in information retrieval systems and decision-making processes.

According to [145], in the context of the Internet of Digital Twins (IoDT), frequent data synchronization interactions between physical entities and digital twins, as well as intensive data exchanges among twins, are essential. These interactions impose significant demands on communication systems, necessitating low-latency and low-overhead communication

capabilities. SC emerges as a promising solution in this context, offering the potential for ultra-low latency semantic transmissions in both intra-twin and inter-twin communications. By leveraging SC, the IoDT can optimize data exchange, transmitting only relevant and necessary semantic content. However, it is crucial to be aware of the threat posed by semantic data or knowledge poisoning attacks. These attacks can occur during interactions between twins and physical entities or between twins themselves. Malicious entities intentionally inject poisoned data samples into raw data or KB with the objective of manipulating model training. Data poisoning mostly occurs at the transmitter end, where these malicious entities utilize contaminated datasets to degrade the performance of DNNs.

The use of KG in recommendation systems is one of the SC applications, where the objective is to provide optimal recommendations to users automatically. KG have proven instrumental in enhancing the explainability of these recommendations, offering insights into the reasoning behind the system's suggestions [146]. This has led to an increased importance for KG-based recommendation systems in real-world scenarios such as music, film, and online shopping domains. These systems have demonstrated improved recommendation accuracy by leveraging KG as auxiliary information. However, despite their numerous advantages, KG is not without vulnerabilities. They are susceptible to poisoning attacks. These attacks aim to manipulate the recommendations by contaminating the KG with fake links, specifically to enhance the visibility of certain products. By injecting fake links into the KG, the attacker can manipulate the system to improve the ranking of specific items in recommendation lists [147].

In the context of multimodal multi-user SC, multiple users collaborate to achieve a common intelligent task, making it particularly suitable for emerging autonomous scenarios in our daily lives. For instance, human activity recognition (HAR) in smart healthcare, where complementary information is collected from various sensors [148]. However, HAR systems often encounter challenges, especially when relying on data from untrusted users. Attackers may exploit this vulnerability by manipulating sensor readings to contaminate the training set. One such attack is the label flipping data poisoning attack, where the labels of sensor readings are maliciously changed during data collection. The presence of high noise and uncertainty in the sensing environment exacerbates the severity of this threat, potentially leading to erroneous outcomes and compromising the reliability of the HAR system [149].

Furthermore, audio intelligence systems often rely on a large corpus of training samples, which can lead them to utilize third-party resources. This reliance on external data sources introduces vulnerabilities to data poisoning attacks and backdoor attacks. Ge *et al*. specifically addressed the vulnerability of audio intelligence systems to these types of attacks. Audio intelligence systems require a substantial amount of training samples, costly computational resources, and expert knowledge, which may be challenging for individuals with limited means to obtain. Consequently, users may resort to utilizing third-party resources, which can expose them to potential data poisoning attacks [150].

2) Knowledge/Data Tampering Attacks: Knowledge/data tampering attacks involve the intentional and malicious act of altering or manipulating data with the aim of deceiving or causing harm. In such attacks, an attacker may modify, forge, replace, or delete data to achieve their desired outcome. This can involve changing the values of data elements, altering records, injecting false information, or removing important data points. The attacker's goal is to manipulate the data in a way that benefits them or serves their malicious intent. Knowledge/data tampering attacks pose a threat within the life cycle of digital twin services in the Internet of Digital Twins (IoDT) [145]. Attackers have the ability to manipulate the semantic data stream and the exchanged interests or content. Moreover, the interpretation of semantic data relies on the receiver's KB, leading to potential discrepancies in the information received by receivers with different background knowledge. This variation can result in physical entities or twins being unaware of modifications made to the semantic data stream during intra/inter-twin interactions.

In recent times, the concept of Integrated SC and AI-Generated Content has garnered considerable interest [151]. This approach involves the seamless transfer of semantic information derived from user inputs, which AI systems use to generate digital content and render immersive graphics within the Metaverse. By incorporating Integrated SC and AI-Generated Content, users can expect enhanced interactions and experiences, as AI-driven content creation opens new frontiers of creativity and interactivity. However, this virtual shared space, merging elements of the physical world with persistent virtual reality, faces significant threats from knowledge/data tampering attacks. Attackers can manipulate data within Metaverse data services by removing specific data to disrupt operations, replacing authentic data with false information, forging data to deceive users, or modifying existing data to achieve malicious objectives. These actions pose substantial risks, leading to inaccuracies, misinformation, and potential harm to users and entities within the Metaverse [152].

Efficient communication is crucial for real-time decision-making in applications like smart grids and networked control systems. In smart grids, real-time decision-making responds promptly to changes in electricity demand and supply, preventing power outages and minimizing energy waste. However, these applications generate vast amounts of data, leading to network bottlenecks. SC offers a solution by reducing the volume of generated and transmitted data [153]. Nevertheless, [154] highlights the vulnerability of IoT-based smart grids to cyber attacks. Interconnected devices create potential entry points for unauthorized access by malicious actors. Data integrity attacks pose significant threats, disrupting grid state estimation and jeopardizing operation and stability. The paper [154] focuses on a zero-parameter-information data integrity attack, exploiting the smart grid's topology vulnerability, allowing stealthy data tampering without knowledge of branch parameters.

C. Transmission Channel Security

The transmission channel is the physical medium through which data is transmitted between the sender and receiver. The

security of the transmission channel is important to ensure that the data is not intercepted or modified by unauthorized parties.

1) Eavesdropping Attacks: Transmitting confidential data poses challenges, especially with numerous low-cost and low-complexity devices. A.D. Wyner's study on a three-terminal wireless channel involving legitimate users Alice and Bob, and an eavesdropper Eve, revealed channel imperfections [155]. An eavesdropping attack occurs when an attacker intercepts communication between a transmitter and receiver to understand the exchanged messages without disrupting the system.

In conventional systems, eavesdroppers aim to obtain the source message sent by the transmitter. In SC systems, the transmission involves semantic information that requires decoding to retrieve the original message. This means eavesdroppers might intercept semantic information but still struggle to obtain the desired content. The paper [17] discusses scenarios where eavesdroppers may succeed in intercepting but fail in decoding information due to differences in background knowledge, disparities in task objectives, or encryption mechanisms in the semantic encoding model.

While conventional communication systems posed difficulties for eavesdroppers in extracting privacy information from noisy channels, SC offers improved efficiency and accuracy, particularly in low SNR scenarios [156]. Unfortunately, this efficiency also presents opportunities for potential eavesdroppers as they can decipher semantic information even in highly noisy channels. For instance, even with poor channel conditions, eavesdroppers can exploit shared decoders to decipher semantic information. Furthermore, semantic information could reveal the actual distribution of user data to some extent, potentially compromising user privacy [145].

2) Physical Layer Adversarial Attacks: The focus of end-to-end SC, as emphasized in [157], is not solely on fully recovering the transmitted message, but rather on empowering the receiver to comprehend the intended meaning and take appropriate actions within the relevant context. With this in mind, the authors of [158] argue that robust interpretation of conveyed semantics at the receiver side is crucial for the success of SC systems. However, achieving robust semantic interpretation is challenging due to the susceptibility of end-to-end SC to physical adversarial attacks, as discussed in [159]. These attacks exploit the vulnerability of wireless channels and the fragility of DNNs. To address these challenges, the paper [158] introduced a framework called MobilrSC, which focuses on computation and memory efficiency in wireless environments for DL-based SC systems. Additionally, the authors proposed a physical-layer adversarial perturbation generator named SemAdv. This generator aimed to craft semantic adversaries over the air, employing imperceptible, input-agnostic, and controllable criteria. These attacks are specifically tailored to distort the receiver's understanding and lead to incorrect decision-making within SC systems.

Adversarial attacks can be broadly categorized into two primary types: white-box attacks and black-box attacks. In white-box attacks, the attacker has complete knowledge of the target model, including its parameters, architecture, and training data. This deep understanding empowers the attacker to strategically design adversarial examples that effectively deceive the target model. Conversely, black-box attacks occur when the attacker has limited or no knowledge about the target model. Consequently, the attacker must rely on indirect methods, such as observing the model's output on known inputs, to craft adversarial examples. Black-box attacks are generally more challenging to execute compared to white-box attacks due to the limited information available to the attacker. Nevertheless, they are considered more realistic in real-world scenarios as attackers typically lack access to the internal details of the target model [126].

In [160], Li *et al*. introduced SemBAT, an approach for generating black-box adversarial attacks in DL-based SC systems at the physical layer. The proposed method involved training a surrogate encoder using gradient estimation and data augmentation techniques based on the Jacobian matrix. By employing the particle optimization algorithm, SemBAT generated adversarial perturbations. These perturbations were introduced as noise during the transmission of representations through the wireless channel. The experimental results highlighted the remarkable effectiveness of SemBAT in significantly reducing the classification accuracy of the SC system. Importantly, these adversarial perturbations remained imperceptible to human observers, as evidenced by image quality metrics.

Moreover, the introduction of perturbations into the embedding layer via the semantic channel can give rise to what is known as semantic noise. This kind of noise has the potential to distort the interpretation of transmitted data, leading to erroneous conclusions or decisions [137]. A similar scenario unfolds within the vehicular metaverse, where vehicles rely on semantic information for communication. The pivotal role of the SC module is to facilitate the exchange of meaningful semantic data between vehicles [161]. However, this system is susceptible to adversarial attacks involving semantic noise. In these attacks, malicious actors strategically insert disruptions into transmitted data in order to fool the communication module. These disruptions are carefully crafted to minimally alter the data position in the semantic space. Despite their subtlety, they have the power to trigger a misinterpretation of information, thereby potentially prompting vehicles to make incorrect decisions.

3) Man-In-The-Middle Attacks: A man-in-the-middle attack (MITM) occurs when an attacker secretly intercepts and potentially alters communications between two or more parties who believe they are directly communicating with each other, with the attacker inserting themselves in the middle. This enables the attacker to read, modify, or even discard the semantic information being transmitted. The impact of a MITM attack on a SC system depends on the specific application. However, generally, a successful MITM attack can severely compromise the security and reliability of the system. For instance, in a remote healthcare monitoring system, a MITM attack could disrupt operations and prevent alarms from being raised, potentially jeopardizing patient care [162]. Similarly, if a MITM attack succeeds on a system controlling an autonomous vehicle, the attacker could take control of the vehicle, leading to a serious accident [163].

The combination of blockchain and SC enables a decentralized and efficient exchange of semantic information among

unfamiliar participants in the Metaverse, ensuring enhanced security. SC reduces the burden of communication and storage for large data sets by facilitating the exchange and processing of semantic information. Meanwhile, blockchain technology establishes trust and safeguards against manipulations and false modifications by attackers or third parties [164]. However, this integration faces the challenge of ensuring robust data security. Effective semantic data sharing relies on interactions between unidentified virtual service providers and edge devices in untrusted environments, facilitated by the blockchain. Before being uploaded to the blockchain, extracted semantic data may undergo manipulation to display similar descriptors (semantic similarities) while conveying different meanings. For instance, a digital twin service provider that seeks images of snowy mountains might receive manipulated images closely mimicking snowy mountains [17]. If digital twin service providers fail to detect such deceitful tampering, their database could be compromised. This form of attack, known as a semantic data poisoning attack, can also be utilized in MITM attacks, where an intermediary node with malicious intent intercepts wireless communication channels and replaces transmitted images without altering the underlying semantic information.

4) Jamming Attacks: A jamming attack involves deliberately disrupting a wireless communication signal to hinder legitimate users from accessing the network. In the context of SC, which focuses on meaningful data exchange, jamming poses new challenges. Unlike traditional channel jamming methods, semantic jamming aims to degrade the quality of the semantic content recovered by the receiver. Attackers must employ more effective jamming techniques to reduce the semantic consistency between the decoded result and the original data. For instance, the authors in [165] proposed an intelligent jamming framework by using a game strategy like the GAN to improve the performance of the semantic jammer.

D. Lessons and Summary

Section IV discusses the security concerns related to SC systems, introducing new security and privacy challenges. One key concern is the vulnerability of the encoder and decoder, which use machine learning algorithms to learn patterns and relationships between raw data and its semantic representation. These models are susceptible to various attacks, such as poisoning attacks that degrade their performance. Additionally, the KB storing semantic representations and relationships is at risk of unauthorized access or tampering, compromising data integrity and confidentiality. The transmission channel is another area of concern, requiring encryption, access control, secure storage, and transmission protocols to protect semantic data. This section highlights these vulnerabilities and proposes mitigation strategies to ensure the reliable and secure operation of SC systems.

V. Countermeasures of Security and Privacy in SC

Building upon the highlighted vulnerabilities, this section underscores the critical need to ensure robust security and protect privacy in SC systems. In this section, we provide the existing defense strategies employed to counter potential threats within the realm of SC systems.

A. Adversarial Training

Adversarial training is a technique used in ML to improve the robustness of a model against adversarial attacks. Adversarial attacks intentionally manipulate input data to mislead the model's output. Adversarial training involves adding adversarial examples to the training data to make the model more resilient to such attacks. The model is trained on both clean and adversarial examples, which helps it recognize and reject adversarial examples. According to [166], Chen *et al*. proposed De-Pois, an attack-agnostic defense against poisoning attacks in ML. The defense strategy involved training a mimic model to imitate the behavior of the target model trained on clean samples. GANs were used to facilitate informative training data augmentation and mimic model construction. By comparing the prediction differences between the mimic model and the target model, De-Pois was able to distinguish the poisoned samples from clean ones, without explicit knowledge of any ML algorithms or types of poisoning attacks.

Similarly, in [167], adversarial training poison immunity was proposed to defend against data poisoning attacks. The method created poisons during training and injected them into training batches to desensitize networks to the effects of such attacks. The poisons were generated by adding small perturbations to a subset of the training data, which were then used to train the model. This process helped the model learn to ignore the poisons and focus on the true data. The defense mechanism was evaluated in different scenarios and was shown to withstand adaptive attacks and generalize to diverse threat models.

To prevent semantic noise from influencing SC systems, the authors in [137] proposed a robust DL-enabled SC system that uses a calibrated self-attention mechanism and adversarial training to tackle semantic noise in text transmission. The calibrated self-attention mechanism helps focus on important parts of the input text, while adversarial training increases the system's robustness against adversarial semantic noise. The system models the transmitter and receiver as neural networks, using a loss function for training. The input text is passed through a one-hot encoder and embedding layer to generate an embedding vector, and a BERT score measures the similarity between the transmitted and reconstructed text. This system shows remarkable performance in dealing with semantic noise under different SNRs compared to baseline models. Similarly, Hu *et al*. proposed a framework for robust end-to-end SC systems to combat semantic noise [138]. This method employs adversarial training with weight perturbation, incorporating samples with semantic noise in the training dataset. It also masks frequently noisy input portions and designs a masked vector quantized-variational autoencoder with a noise-related masking strategy. The system uses a discrete codebook shared by the transmitter and receiver for encoded feature representation. To further enhance robustness, a feature importance module suppresses noise-related and

task-unrelated features, allowing the transmitter to send only the indices of important task-related features. The simulation results demonstrate significant improvements in robustness against semantic noise and a notable reduction in transmission overhead.

In [168], the authors proposed a defense against membership inference attacks on DL models using generative adversarial networks (GANs). The proposed defense aimed to maintain the accuracy of the model while protecting privacy against membership inference attacks. The defense involved training a GAN on sensitive data and using it to generate data for training the actual model. Two different GAN structures with special training techniques were utilized to deal with the image data and the table data, respectively.

In addition, real-time applications are commonly employed in critical scenarios where the impact of an attack can be extremely severe. For instance, in the case of a real-time image recognition system utilized in a self-driving car, a successful adversarial attack could lead to disastrous consequences such as a potential collision. In [169], the authors specifically addressed the challenges faced in image recognition applications where the ground-truth of incoming images is unknown, rendering the computation and validation of classifier accuracy impossible. To counter such attacks, the authors proposed a privacy-preserving framework that defends black box classifiers by utilizing an ensemble of iterative adversarial image purifiers. The proposed approach effectively transformed a single-step black box adversarial defense into an iterative defense strategy. Additionally, the paper introduced three innovative privacy-preserving knowledge distillation techniques. These approaches leveraged prior meta-information from diverse datasets to emulate the performance of the black box classifier. Notably, the paper established the existence of an optimal distribution for purified images, which can reach a theoretical lower bound. Beyond this threshold, the image can no longer be purified.

While images are typically represented in two dimensions, audio signals exist as one-dimensional time series data. Unlike images, which are often analyzed as a whole or in patches without strict order constraints, audio signals must be sequentially examined in chronological order. Although audio signals can be transformed into two-dimensional time-frequency representations, the axes of time and frequency differ fundamentally from the horizontal and vertical axes of an image. These unique properties necessitate audio-specific transformations, which can be performed in either the time or frequency domain. In [170], the authors explored the vulnerability of speaker recognition systems to adversarial attacks and proposed defenses based on transformations and adversarial training to enhance their security. The authors presented 22 diverse transformations and evaluated their effectiveness against seven recent adversarial attacks targeting speaker recognition systems. They assesed the resilience of these transformations against adaptive attacks and measure their efficacy when combined with adversarial training. The proposed approach involved a novel feature-level transformation combined with adversarial training, which proves to be more effective compared to sole adversarial training in a complete white-box setting. By leveraging these defense strategies, the speaker recognition systems demonstrated improved robustness against adversarial attacks in the audio domain.

B. Data Denoising

Data denoising is the process of removing unwanted noise from a dataset. Noise is irrelevant or random variations that can distort the actual information. Liu *et al*. proposed a countermeasure for poisoning attacks on DNNs used in human-computer interactions called Data Washing [171]. The Data Washing algorithm is based on a denoising autoencoder. The data are first passed through a denoising autoencoder. A small amount of Gaussian noise is then added to the data and the data are then passed through the autoencoder once again to obtain the restored data. The algorithm removes the malicious signal added by the attacker and provides effective protection against the attacker.

In case of textual context, the paper [121] introduced a defense technique called DeBITE, which was designed to counter the BITE backdoor attack in textual contexts. DeBITE employed a potential trigger word removal approach, which involved the identification and elimination of trigger words from the training data. The method operated by initially identifying potential trigger words that exhibit a strong correlation with the target label. Subsequently, these identified words were removed from the training data, and the model was retrained using the cleaned dataset.

C. Covert Communications

Covert communication networks aim to conceal transmitted signals or semantic information, making it challenging for attackers or wardens to detect or decode them. This is accomplished by introducing randomness or noise into the transmission, effectively camouflaging the semantic information. Covert communication networks are commonly utilized in situations requiring secrecy or confidentiality, such as military operations, espionage, or secure communication channels. By disguising signals within noise, these systems aim to avoid detection by unauthorized individuals monitoring the communication channel. In [172], the authors proposed a framework for covert SC for image transmission over wireless networks. In this framework, devices extracted and selectively transmitted semantic information of image data to a base station. The semantic information consisted of the objects in the image and a set of attributes of each object. A warden selected a device to detect and eavesdrop on semantic information. To ensure the security of SC, a jammer acted as the defender and transmitted jamming signals to the vulnerable device. The proposed algorithm enabled each device and the jammer to cooperatively discover the vulnerable devices as well as find the semantic information transmission and power control policies that maximize the performance of the covert SC system.

D. Encryption

This method encrypts the semantic information transmitted over the channel, making it more difficult for an attacker to eavesdrop on or modify it. Encryption techniques can make it difficult for an attacker to decode the semantic information, even if they detect it. The authors in [173] introduced DeepJSCEC, a secure wireless image transmission scheme building upon DeepJSCC. DeepJSCEC used mapping techniques to recover input with minimal distortion despite channel noise. However, it was susceptible to eavesdropping due to the inherent correlation between the source sample and channel input. The method in [174] aimed to solve this security problem using symmetric encryption and an adversarial training scheme to maintain encryption feasibility and security. The proposed system consisted of an encryptor and an attacker. The encryptor encrypted the semantic information using a symmetric encryption algorithm and sent it to the receiver. It was trained using an adversarial training scheme to balance the utility and confidentiality of the encrypted message. The attacker intercepted the encrypted message and tried to reconstruct the semantic information directly using a semantic attacker, trained similarly to minimize a defined loss function. Experimental results showed that the proposed method achieved better performance in terms of security and utility compared to existing methods.

Moreover, the authors in [175] identified a phenomenon in SC systems due to the flexibility of semantic transmission, which did not require strict matching between decoding and encoding sequences. This led to variations in words and fixed sentences, known as semantic drifts. To leverage this randomness, they proposed SemKey, a physical layer key generation scheme for securing DL-based SC systems. SemKey used the random features of SC to create a switch sequence with varying characteristics for the reconfigurable intelligent surface-assisted channel. By using the parallel factor-based channel detection method, they performed channel detection in the presence of reconfigurable intelligent surface assistance. This approach significantly improved the rate of secret key generation.

Ensuring randomness is crucial for generating secure secret keys. In [176], the authors proposed a physical layer semantic encryption scheme to enhance the security of DL-based SC systems. The proposed method used the randomness in BLEU scores from machine translation. By feeding the weighted sum of these scores into a hash function, semantic keys were generated, producing secure and unpredictable keys for semantic encryption. Additionally, they introduced a semantic obfuscation mechanism involving subcarrier obfuscation with dynamic dummy data insertion. Experimental results demonstrated the effectiveness of the proposed method, particularly in static wireless environments.

E. Jamming Resistant

To counter intelligent jamming attacks effectively, it is essential to employ an intelligent receiver capable of maintaining semantic consistency in the presence of jamming. The study in [165] focuses on the receiver's ability to accurately decode semantic information even when faced with semantic jamming attacks. The proposed framework employs a game model that optimizes both the semantic jammer and the receiver, enhancing the robustness of the SC system. This game strategy, inspired by GAN), aims to significantly improve the receiver's performance.

The authors of [177] argued that traditional anti-jamming methods, which focus on detecting and mitigating jamming attacks by identifying the jamming policy and applying countermeasures, might be inadequate against sophisticated jamming attacks employing dynamic or adaptive jamming policies. To address this, they proposed an anti-jamming defense method that involves recognizing the jamming policy and selecting appropriate countermeasures. The method leverages RNN to handle the sequential nature of interactions between the user and the jammer. By employing RNN, the authors aimed to capture the evolving dynamics of jamming attacks and develop effective defense strategies. They also considered scenarios with multiple jammers using distinct policies, proposing methods to estimate future behavior using RNN for proactive defense measures. This comprehensive framework aims to enhance the effectiveness of anti-jamming strategies against various types of jamming attacks, including dynamic or adaptive policies.

SC also enabled a macro base station to interpret various traffic signs, allowing it to make decisions for connected and autonomous vehicles [178]. While the interconnected feature of intelligent vehicles is typically advantageous for an Intelligent Transportation System, it also exposes the vehicular network to the risks of illegal jamming and eavesdropping. To address these risks, the authors in [179] proposed a technique to enhance the security performance of connected and autonomous vehicle networks against illegal eavesdropping and jamming interference. The proposed technique utilizes distributed Kalman filtering and DRL techniques to improve anti-eavesdropping communication capacity. A distributed Kalman filtering algorithm was developed to more accurately track attackers by sharing state estimates among adjacent nodes. The authors formulated a design problem to control transmission power and select communication channels while ensuring the communication quality requirements of authorized vehicular users. They developed a hierarchical DQN-based architecture to design anti-eavesdropping power control and potential channel selection policies. Initially, the optimal power control scheme, without prior information about eavesdropping behavior, could be quickly achieved. After assessing the system's secrecy rate, the channel selection process was performed if necessary.

F. Authentication

Authentication is the process of verifying the identity of a user or device, acting as the first line of defense against unauthorized access. This is achieved by requiring evidence such as something the user knows, has, or is. The utilization of SC in the Metaverse introduces significant security concerns for AI-Generated Content. Attackers exploit this by sending deceptive semantic data that closely resembles authentic information but contains different content, aiming to disrupt Metaverse services. The challenge lies in virtual service providers

distinguishing between adversarial and genuine semantic data, as differences may not be easily detectable. The nature of virtual networks in the Metaverse complicates detecting and preventing semantic data mutations due to the distributed nature of virtual service providers and edge devices. To address these concerns, Lin *et al*. proposed a semantic defense scheme leveraging blockchain and zero-knowledge proofs to distinguish between adversarial and authentic semantic data and verify the authenticity of semantic data transformations [164]. This scheme ensured properties such as completeness, soundness, and zero-knowledge. The process involved transforming semantic data into a commitment using a Pedersen commitment scheme, storing it on the blockchain, generating a zero-knowledge proof for authenticity, and verifying the proof with the provided key. If the proof was valid, the semantic data transformation was considered authentic.

The global shift from fossil fuels to electric vehicles (EVs) necessitates a robust charging infrastructure. A user-friendly, cost-effective charging network is crucial for widespread EV adoption. Charging methods can be static or dynamic, with dynamic charging facing collaboration challenges among EVs, road infrastructure, and charging stations (CSs). EVs use the dedicated short-range network, based on IEEE 802.11p, to communicate with CS infrastructure. Roadside units (RSUs) enable EVs to connect to CSs through backbone networks. However, communication channels between EVs, RSUs, fog servers, and company charging servers are insecure, making them susceptible to eavesdropping, MITM, and jamming attacks [180].

The Internet of Vehicles (IoV) serves as the central infrastructure for providing advanced services to connected vehicles and users, enhancing transportation efficiency and security. IoV-enabled traffic management systems optimize traffic flow and improve safety by leveraging real-time information from EVs. These systems identify traffic bottlenecks and make informed decisions on traffic rerouting, proactively managing traffic conditions and mitigating congestion-related issues. However, the explosive growth in emerging applications and services within the IoV poses a significant challenge of spectrum scarcity, as mobile data traffic between connected vehicles and RSUs continues to rise. To address this, Xu *et al*. proposed a cooperative semantic-aware architecture to reduce data traffic in the IoV. Their approach involved conveying essential semantics from collaborated users to servers. A cooperative semantic feature recovery approach utilizing a Joint Source-Channel decoder enabled the recovery of semantic features from multiple cameras for identification purposes [181].

Given the insecure nature of communication channels in such systems, IoV-enabled CSs can integrate with the IoV, enabling real-time communication between EVs and CSs. These stations provide EVs with up-to-date information regarding charging availability and pricing. An authenticated key agreement protocol for dynamic charging of EVs was proposed in [180]. This protocol ensured confidentiality and authentication during the charging process. The EV initiated a charging request to the CS, which responded with a challenge. The EV provided a valid response along with its identity. The CS verified the response and shared a session key with the EV. Throughout the charging session, the EV and CS used the session key for secure message encryption and decryption. The protocol guaranteed mutual partnership and session identifier sharing, incorporating a mechanism to prevent double spending by checking the key against a revocation list.

G. Threat Detection

Threat detection is detecting abnormal activities in data to find potential security risks or attacks. In [171], the authors propose an integrated detection algorithm to detect various types of attacks. The algorithm is based on the analysis of the output of the penultimate layer of the model. The algorithm uses a threshold value to determine whether the output of the penultimate layer is normal or abnormal. If the output is abnormal, the algorithm will classify the data as attacks. The integrated detection algorithm provides an accurate means of detecting data sets that contain abnormal data and thus provides effective protection against attacks. In particular, the proposed detection method is applicable to different DNN models.

In [182], the authors proposed a novel approach to counteracting backdoor attacks on DNNs. The method introduced the concept of a detrigger autoencoder, which effectively removed the trigger embedded within backdoor samples. Using this technique, the proposed method detected backdoor samples by observing the subsequent changes in the classification results. Furthermore, Xiang *et al*. addressed the challenge of post-training detection in DNN image classifiers, specifically focusing on scenarios where the defender lacked access to the poisoned training set [183]. Instead, they only had access to the trained classifier itself and clean examples from the classification domain. This situation was particularly relevant when considering applications such as widely shared phone apps, where the classifier's integrity affected numerous users. The authors proposed a purely unsupervised anomaly detection defense mechanism against subtle backdoor attacks. This defense not only identified whether the trained DNN had been compromised, but also inferred the source and target classes involved in the attack, while estimating the underlying backdoor pattern. By offering a robust solution that operated without access to the training set, the proposed approach demonstrated promising potential in countering backdoor attacks on trained classifiers.

Wang *et al*. proposed a defense mechanism known as feature manipulation defense [184]. This mechanism detects and cleanses adversarial examples efficiently and interpretably. The defense approach is based on the observation that the classification outcome of a regular image typically remains unchanged despite non-significant intrinsic feature alterations, whereas adversarial examples are highly sensitive to such modifications. To enable feature manipulation, the authors employed a Combo-variational autoencoder (Combo-VAE) to acquire disentangled latent codes revealing semantic features. The resistance of the classification outcome to morphological changes, generated by varying and reconstructing these latent codes, is leveraged to identify suspicious inputs. Additionally,

the Combo-VAE is enhanced to clean adversarial examples by considering both class-shared and class-unique features, resulting in high-quality purified examples.

H. Differential Privacy

DL models are often trained on large datasets that contain sensitive information about individuals. Differential privacy (DP) can be used to prevent MI attacks by adding noise to the model's output, making it more difficult for an attacker to reconstruct the input data used to train the model. DP serves the fundamental purpose of safeguarding individual privacy by enabling data analysis without revealing sensitive personal information. This approach ensures data analysis can proceed while preserving individuals' confidentiality. Moreover, DP maintains the principle that the inclusion or exclusion of any individual's data in the dataset minimally influences the results of the published query. DP has gained significant attention and importance in the era of big data, where large amounts of personal data are collected and analyzed.

The work in [185] discussed the need for transparency in ML models to increase trust, ensure accountability, and scrutinize fairness. However, organizations may opt out of transparency to protect individuals' privacy. Therefore, there is a demand for transparency models that consider both privacy and security risks. The authors introduced a technique that complements DP to ensure model transparency and accuracy while being robust against MI attacks.

Similarly, in [186], Ye *et al*. proposed a time-efficient defense method against both membership inference and MI attacks. The method required only one parameter, the privacy budget, to be tuned. The privacy budget is a key parameter in differential privacy, controlling the amount of noise added to the confidence score vectors to protect against privacy attacks. The paper theoretically demonstrated how to tune the privacy budget to defend against both types of attacks while controlling the utility loss of confidence score vectors.

I. Data Protection

Data protection involves implementing measures and strategies to safeguard data from unauthorized modifications, alterations, or tampering attempts. Maintaining the accuracy, consistency, and unaltered state of information within a KB for SC systems is crucial. This ensures that the KB serves as a reliable and dependable source of information for users and applications. By preserving data integrity, SC systems can effectively support decision-making processes, and facilitate seamless interactions based on trustworthy information. Consequently, the study presented in [187] introduced VBlock, a blockchain-based tamper-proof data protection model for IoV networks. VBlock leveraged the blockchain to ensure data immutability and resistance to tampering, effectively addressing concerns related to unauthorized alterations of outsourced vehicular data in smart city management and enhancement. VBlock introduced an innovative, collusion-resistant model for outsourcing data to cloud storage, which helped maintain the network's tamper-proof nature while ensuring robust data provenance and auditing capabilities. Moreover, it incorporated a key revocation mechanism that enhanced network security against malicious nodes. VBlock was built upon a Hyperledger Fabric blockchain, known for its heightened security and privacy achieved by restricting access to recognized nodes. The proposed model demonstrated substantial security assurances coupled with high efficiency, rendering it applicable and feasible within the IoV environment.

J. Quantum Key Distribution

Quantum Key Distribution (QKD) is a pivotal countermeasure for enhancing the security of SC systems. QKD employs the principles of quantum mechanics to securely distribute cryptographic keys between parties, ensuring that any interception attempts are detected [189]–[192]. This method leverages the fundamental properties of quantum mechanics, such as Heisenberg's uncertainty principle and the no-cloning theorem, to provide provable security against eavesdropping. In SC systems, the transmission of semantic information involves sensitive data that must be protected from unauthorized access. Traditional encryption methods, while effective to some extent, are vulnerable to the increasing computational power of attackers, especially with the advent of quantum computing. QKD addresses this challenge by enabling the secure generation and distribution of cryptographic keys that can be used for one-time pad encryption, which is information-theoretically secure.

The implementation of QKD in SC systems involves the use of quantum channels to transmit quantum states that encode the cryptographic keys. Any attempt to intercept these keys alters their state, alerting the communicating parties to the presence of an eavesdropper. Additionally, classical channels are used for key reconciliation and error correction processes, ensuring that the keys shared between the sender and receiver are identical and error-free. QKD offers several advantages in SC systems. First, it ensures that any interception attempts are immediately detected, providing a level of security unattainable with classical cryptographic methods. Second, QKD allows for the generation of new keys for each communication session, reducing the risk of key reuse and potential compromise. Third, with the development of quantum computers, many traditional cryptographic methods may become obsolete, but QKD is inherently resistant to such advancements, ensuring long-term security.

A study presented in [188] introduced a QKD-secured semantic information communication (QKD-SIC) system for intelligence-native 6G networks. This system connects edge devices via quantum channels to encrypt and securely transmit semantic information. The proposed QKD-SIC system addresses the challenge of unpredictable semantic information generation by edge devices and optimizes QKD resources through a two-stage stochastic optimization model. This model ensures efficient resource allocation and cost reduction, demonstrating the feasibility and effectiveness of QKD in protecting SC system. The QKD-SIC system leverages a global resource pool created by cooperative QKD service providers. By sharing QKD and key management wavelengths, the system efficiently utilizes resources, meeting

TABLE VII
SUMMARY OF THE EXISTING SECURITY AND PRIVACY ATTACKS AND THE COUNTERMEASURES IN SC SYSTEMS

Target	Attack Name	Key Features	Countermeasures
Encoder/Decoder	Poisoning Attacks [107], [108], [111], [112], [114], [115]	Add the malicious or fraudulent data to the training dataset.	Adversarial Training [166], [167], Data Denoising [171], Threat Detection [171]
	Backdoor Attacks [107], [116]–[120]	Introduce a hidden trigger into the model to cause incorrect predictions when activated.	Data Denoising [121], Threat detection [182], [183]
	Adversarial Example Attacks [107], [123]–[129]	Manipulate the input data in a way that appears innocuous to humans but leads to misclassification or incorrect output from the system.	Adversarial Training [169], [170], Authentication [164], Threat detection [184]
	Model Inversion Attacks [107], [130]–[133]	Use the output of a machine learning model to recover the private dataset that was used to train the model.	Differential Privacy [185], [186]
	Membership Inference Attacks [107], [134]–[136]	Try to infer whether a particular input was part of the training dataset of a machine learning model.	Adversarial Training [168], Differential privacy [186]
	Semantic Noise [137]–[139]	Typos or grammatical errors in the textual context or introduce the imperceptible perturbations to the data.	Adversarial Training [137], [138]
Knowledge Base	Data Poisoning Attacks [107], [145], [147], [149], [150]	Add the malicious or fraudulent data to the knowledge base or dataset.	Adversarial Training [166], [167], Data Denoising [171], Threat detection [171]
	Data Tampering Attacks [145], [152], [154]	Unauthorized modification or alteration of data to deceive, disrupt, or gain advantage.	Data Protection [187]
Transmission Channel	Eavesdropping Attacks [145], [156]	Intercept and listen to the communication channel between a transmitter and receiver.	Covert Communications [172], Encryption [174]–[176], Authentication [180], Quantum Key Distribution [188]
	Physical Layer Adversarial Attacks [137], [158]–[161]	Introduce perturbations as noise during the transmission through the wireless channel.	Adversarial Training [137], [138]
	Man-In-The-Middle Attacks [17], [162], [163]	Secretly intercept and alter the communications between two parties who believe they are directly communicating with each other.	Authentication [180]
	Jamming Attack [165]	Degrade the quality of the semantic content recovered by the receiver.	Jamming Resistant [165], [177], [179], Authentication [180]

the secret-key requirements for semantic information transmission. Additionally, the use of Shapley value from cooperative game theory ensures fair cost-sharing among service providers, further enhancing the system's practicality.

Experimental results from the study indicated a substantial reduction in deployment costs—approximately 40% compared to existing non-cooperative baselines. This cost efficiency, combined with the heightened security provided by QKD, underscores the significant benefits of integrating QKD into SC systems. In conclusion, QKD represents a robust countermeasure for securing semantic communication system. By ensuring secure key distribution and leveraging quantum mechanics' inherent properties, QKD enhances the confidentiality and integrity of semantic information, making it a critical component in the next generation of secure communication technologies.

K. Lessons and Summary

In Section V, we explored various defense strategies to address security and privacy concerns in SC systems. Adversarial training enhances model robustness by training on both clean and adversarial examples, helping models recognize and reject adversarial inputs. Techniques like De-Pois and GAN-based defenses tackle poisoning and membership inference attacks. Data denoising methods like Data Washing and DeBITE counter backdoor attacks and remove noise in textual contexts. Covert communication networks and encryption methods, such

as DeepJSCEC and SemKey, secure semantic information by introducing randomness and making it difficult for attackers to detect or decode transmissions. Jamming resistance strategies use intelligent receivers and anti-jamming methods with RNN and distributed Kalman filtering to enhance security in connected vehicle networks. Authentication ensures user or device identity verification, employing blockchain and zero-knowledge proofs to distinguish between adversarial and genuine semantic data. Threat detection algorithms, like de-trigger autoencoders and unsupervised anomaly detection, identify potential security risks. Differential privacy prevents MI attacks by adding noise to model outputs, safeguarding individual privacy while allowing data analysis. Data protection with VBlock ensures data immutability and resistance to tampering using blockchain technology. Quantum Key Distribution (QKD) enhances SC security by securely distributing cryptographic keys using quantum mechanics, demonstrated in a QKD-secured semantic information communication system for 6G networks. These strategies collectively improve the robustness, confidentiality, and integrity of SC systems.

A summary of existing works for Section IV, V is presented in Table VII.

VI. Open Challenges and Future Research Directions

This Section VI specifies and discusses the challenges and open research issues to spur further investigation of resource management, security and privacy in SC.

A. Generic Semantic Metrics

A generic semantic metric is essential in SC to provide a common standard for evaluating and comparing the accuracy of semantic information across various application scenarios. SC deals with diverse data types such as text, audio, images, and video, each requiring different performance metrics. However, these individual metrics lack uniformity, making comprehensive performance evaluation challenging. A generic semantic metric addresses this issue, offering a unified evaluation standard for system design, analysis, and optimization. The current lack of unified performance assessment metrics is a significant open challenge in the field of SC, hindering system comparison and interoperability.

Several general semantic metrics include the General Quality Index of Semantic Service [193], which compares the task performance of transmitted and received information; Triplet Drop Probability [194], which indicates the probability of specific bit errors; Semantic Mutual Information [195], quantifying semantic-level distortions during compression for downstream AI tasks; Semantic Impact [11], assessing the influence of semantic information on communication outcomes; Communication Symmetry Index [11], measuring the balance of semantic information exchange between communication parties; and Reasoning Capacity [11], evaluating the system's inferential abilities.

Despite these advancements, the journey towards establishing a universal performance metric for SC is still in its infancy. The pressing need for further research to define, refine, and standardize semantic metrics is evident. Developing a comprehensive and well-defined generic performance metric is crucial, as it would not only enhance the efficacy of contemporary communication systems but also pave the way for future research, fostering innovations and advancements in the domain. A concerted effort in this direction would lead to better interoperability, more effective system comparisons, and a stronger foundation for the ongoing evolution of SC technologies.

B. Advanced Learning for Generalizable Intelligence

In SC systems, integrating advanced learning techniques like transfer learning, meta learning, and continual Learning is crucial for achieving generalizable intelligence and enhancing learning efficiency. Transfer learning particularly enhances the effectiveness of transferring knowledge from one task to related tasks, proving invaluable in contexts where devices handle multiple tasks or face limited training data. For instance, Nguyen *et al*. [196] demonstrate optimizing multi-user SC through transfer learning and knowledge distillation, significantly boosting performance for users with varying computing capabilities by facilitating knowledge transfer from high-capacity to low-capacity user models. Similarly, Wu *et al*. [197] introduce a novel transfer learning strategy to guide the training process in object detection with limited labels by leveraging semantic information across tasks, enhancing few-shot detection performance and reducing IoT devices' storage pressures. These studies underscore transfer learning's role in addressing key challenges like multitasking and performance optimization under resource constraints, solidifying its importance in SC research.

Meta learning, or "learning to learn," allows for quick adaptation to new tasks with minimal data, whereas continual Learning focuses on acquiring new information while retaining previously learned knowledge, employing methods like regularization, replay, and parameter isolation. These techniques are essential for developing more robust and adaptable AI systems, promising more effective communication within semantic frameworks. Nonetheless, these areas remain underexplored, presenting significant open challenges and future research directions.

C. Multimodal Semantic Transciver

For 5G/6G applications like the *internet of no things* (metaverse) [198], the data to be transmitted is typically multimodal, encompassing text, voice, images, videos, and more. With the widespread deployment of various sensor devices, multimodal data has become the most important means of information generation in modern society. As a result, a multimodal SC system is highly required to facilitate communication across these multiple modes. Conventional SC systems are designed to handle only one type of unimodal data. If transmitting multimodal data requires using multiple separate unimodal SC systems, each catering to a specific type of data, it implies that each device must deploy multiple SC systems, potentially leading to significant overheads and inefficiencies. Therefore,

integrating these into a single, cohesive multimodal SC system is essential to streamline operations and enhance efficiency.

Considering the principles and challenges of multimodal DL, several open challenges exist in designing a multimodal SC system [199], [200]. First, there is a need for efficient representation learning that reflects the heterogeneity of different modalities. Second, solving the alignment problem is crucial, as it involves identifying connections between modality elements. Third, reasoning over complex, multimodal data requires effective modeling of interactions to compose and infer knowledge. Fourth, the challenge of generating raw modalities that reflect cross-modal interactions, structure, and coherence must be addressed. Lastly, the system must tackle the problems of knowledge transfer and quantification to process and evaluate multimodal data efficiently. Additionally, mitigating the impact of noise during data transmission through noisy wireless channels and considering bandwidth consumption limitations are critical for achieving optimal performance in collaborative intelligence scenarios. These open challenges highlight the necessity for further research and development in the field of multimodal SC systems.

D. Distributed SC Framework

SC system maximizes bandwidth efficiency by transmitting meaningful information, but maintaining this efficiency requires continuous updates to the semantic encoder, leading to significant energy consumption and privacy issues. Updating these encoders is computationally intensive, draining battery life in mobile and edge devices and exposing sensitive data during the process. An efficient distributed learning framework is essential to address these challenges. Federated learning offers a solution by training and updating models across decentralized devices, enhancing privacy and reducing energy burdens without centralizing data. Xie and Qin's study [43] proposed a lightweight ML model for distributed semantic encoders, improving efficiency and privacy. Similarly, Qin *et al*. [201] introduced a comprehensive SemCom framework integrating users and terrestrial base station edge clouds, showcasing federated learning's potential in SC system. These studies underscore federated learning's role in developing efficient distributed frameworks for SC. However, existing frameworks do not fully address the unique challenges of each network domain, necessitating tailored redesigns [202].

However, the existing distributed learning frameworks designed for SC in generic networks [43], [201], [202] need customization to the unique characteristics of various 5G/6G network scenarios. To leverage the full potential of SC in these networks, it is crucial to develop system models tailored to specific use cases and challenges. In telehealth, SC enhances remote diagnoses by transmitting meaning-based information, but challenges like data accuracy, latency, and privacy must be addressed. In smart cities, SC efficiently manages IoT data, improving systems like traffic management through real-time analysis. However, ensuring accurate data integration remains a challenge. Therefore, existing SC frameworks need to be customized to different network scenarios, focusing on optimizing performance while addressing unique operational challenges.

E. Adaptive Resource Management with Dynamic Network

The dynamic nature of network conditions, influenced by countless factors such as fluctuating user numbers and unpredictable interferences, poses a significant challenge in SC. This ever-changing environment underscores the need for adaptive resource management systems. Real-time adaptability is crucial; systems must not only detect changes in network conditions swiftly but also assess network load and performance post-detection. Following detection, these systems must have the agility to adjust resource allocation strategies on-the-fly, optimizing transmission rates, reallocating bandwidth, or dynamically altering transmission protocols. According to Y. Zhu *et al*. [68], adaptive control involves dynamically controlling transmission volume and rate based on the changing network environment. Additionally, Zhang *et al*. [47] emphasize controlling the rate in low SNR environments considering the rate-distortion trade-off. These approaches enhance communication efficiency and reliability, ensuring efficient use of resources and maintaining stability even under dynamic conditions. As we advance in the era of SC, the emphasis on adaptability is essential. The future of effective and meaningful digital communication depends on developing resource management systems capable of handling dynamic network challenges.

F. Multi-user Multi-task based SC System

In the context of multi-user SC, there are several open challenges that need to be addressed. A unified framework to support various tasks with multimodal data is currently lacking [203]. Key challenges include the reduction of inter-user interference, and processing/fusing received semantic information at the receiver for the transmission of multimodal data. Specifically, reducing interference from other users is critical for both single-modal and multimodal communications. Additionally, effectively fusing and processing the multimodal data at the receiver to ensure accurate and efficient communication is a significant challenge that needs to be tackled to advance the field of multi-user SC [203].

Furthermore, in multi-user scenarios, such heterogeneous situations can coexist with traditional bit-based communication and SC. Handling this requires a unified framework. It is crucial to analyze which communication type is superior under various situations. The semi-NOMA scheme proposed in [78], [96] addresses this by splitting the bit stream into two parts: one transmitted with the semantic stream over a shared sub-band and the other over a separate orthogonal sub-band. This approach efficiently utilizes limited bandwidth while maintaining the performance of both communication types. Future research will likely focus on traditional bit and semantic networks, addressing challenges such as communication mode switching, semantic fairness-driven resource allocation, and long-term network optimization.

Another significant challenge is the complexity of resource allocation considering the computing capacity differences among users [196]. Users in SC systems have varying computing capabilities, necessitating efficient management. For instance, supporting both high-performance and low-performance decoders requires training source-channel coders

with each decoder type, a process that is time-consuming and resource-intensive. To address this, a novel training procedure is proposed where the encoder is first trained with the high-performance decoder. With the encoder parameters fixed, the low-performance decoder is subsequently trained. This approach stabilizes and accelerates the training process, making it more efficient.

G. C4 Orchestration in Satellite-Borne Edge Cloud Network

Utilizing SC technology in satellite-borne edge cloud (SEC) for offloading is crucial for next-generation wireless communication systems [202]. SC significantly enhances transmission efficiency by extracting and conveying the semantic meaning of data through machine learning. This optimization is vital in satellite links, which often suffer from high propagation delays. SC helps optimize spectrum efficiency, reduce energy consumption, and protect user privacy, thereby enabling high-quality service and rapid data processing even in remote or disaster-stricken areas.

However, there are several challenges. Real-time updating of semantic coders introduces issues related to the mobility of SEC, low tolerance for service interruptions, energy consumption, and privacy concerns. Existing distributed learning frameworks do not seamlessly apply to SEC networks, necessitating new approaches. While SC reduces communication load, it increases computational load, requiring the development of optimal computational task strategies that consider various operational factors, including access methods, task processing entities, latency, energy consumption, and privacy. Additionally, integrating SC with SEC networks is crucial for enhancing C4 functionalities. This integration requires real-time updating and synchronization of semantic coders, efficient caching mechanisms, robust control algorithms, and secure, reliable connectivity. Addressing these challenges is essential for the successful implementation of C4 functionalities in SEC networks enhanced with SC.

H. Redefinition of Security Metrics for SC

The paper [17] argue that traditional wireless communication security techniques, designed for bit transmission, fall short when applied to the Semantic Internet of Things (SIoT), which prioritizes the transmission of semantic information. One key reason of this issue lies in the absence of security performance indicators tailored for SC. The lack of new security performance indicators is considered a significant open fundamental challenge. Traditional security performance indicators for bit transmission cannot be directly applied to semantic information transmission. Therefore, there is a need to develop new indicators that can accurately measure the security of SC systems.

The paper [17] propose new security performance indicators which captures the unique characteristics and requirements of SC in the SIoT. One proposed indicator is the semantic secrecy outage probability, which describes the probability that an eavesdropper successfully obtains the semantic information sent by the transmitter and accurately performs the semantic decoding. Another indicator is the detection failure probability, which describes the probability that no transmission activity is detected by a warden during the transmission time of the data. By defining and analyzing these new security performance indicators, researchers can better understand the vulnerabilities and risks associated with SC. This enriched understanding paves the way for the development of robust security techniques and mechanisms, finely tuned to meet the specific challenges and demands of SC.

However, the research of developing new security indicator is notably under-researched, and the proposed indicators are in their initial stages, requiring rigorous scrutiny and comprehensive validation to prove their effectiveness in real-world scenarios. Moreover, it is imperative for more researchers to engage in this endeavor, contributing to the development of new indicators. Such collaborative efforts are essential in fostering a secure SC environment, especially in the evolving landscape of SIoT.

I. Robust Semantic Transceiver Design against Semantic Noise

In the Shannon and Weaver model, while communication at the first level deals with physical noise, it is crucial to define and characterize Semantic Noise at the second level, which occurs in SC. Unlike well-discussed physical noise in wireless channels, the study and modeling of semantic noise in wireless communication remain underdeveloped. Semantic noise leads to misunderstandings and decoding errors, creating a mismatch between the intended and reconstructed semantic meanings at the receiver. This noise can arise during various stages: semantic encoding, data transmission, and decoding [138]. During semantic encoding, it manifests as a mismatch due to the encoder's representational limitations. In data transmission, channel fading and malicious signals introduce semantic noise. Decoding stage noise stems from misinterpretations and ambiguous symbol representations. Semantic noise varies across different sources, such as text and images [138]. In text, it involves semantic ambiguity from slight word changes, while in images, adversarial samples cause noise without perceptible changes to humans. To address these challenges, it is essential to develop a robust SC system, like DeepSC, that jointly considers physical and semantic noise, effectively combating semantic noise impacts with minimal transmission overhead.

J. Quantum Key Distribution in SC

SC systems often deal with sensitive data, such as personal and financial information. Resource allocation algorithms need to ensure that security is maintained throughout the resource allocation process. The research gap that led to the emergence of the ultimate research question is the lack of comprehensive studies on the integration of SC and quantum key distribution (QKD) in the context of 6G communications. While there have been separate studies on SC and QKD, there is a need to explore the potential synergies and challenges of combining these two technologies to enhance the security and efficiency of communication systems in the future. This research question aims to address this gap by investigating the integration of

SC and QKD in the context of 6G communications and exploring the optimal resource allocation and routing strategies to achieve secure and efficient communication.

K. Joint Consideration of Resource Allocation and Security

The rapidly evolving landscape of the IoT and smart cities highlights the importance of jointly considering resource allocation and security. This integrated approach promises optimal resource utilization while ensuring system resilience and robustness. The necessity and benefits of this approach are evident in the IoT ecosystem, where focusing on both resource allocation and security is not just beneficial but essential [204]. Key reasons include optimal resource utilization, balancing the trade-off between performance and security, and enhancing system resilience and robustness. By integrating security requirements into resource allocation, efficient and effective resource utilization can be achieved, ensuring sufficient resources for security measures. This balance between performance requirements (e.g., low latency, high throughput) and security needs (e.g., encryption, access control) is critical for achieving desired levels of both. Furthermore, integrating security into resource allocation enhances system resilience and robustness, enabling better recovery from security incidents or attacks.

In smart cities, the joint consideration of resource allocation and security is crucial for efficient and secure communication in the Internet of Digital Twins [145], [205]. Resource allocation involves the efficient use of communication resources such as bandwidth, power, and computing resources to support secure SC. Security measures are necessary to protect the communication system from threats and attacks, ensuring confidentiality, integrity, and availability of transmitted information. Technologies like blockchain, DL, and MEC play significant roles in achieving these objectives. Despite the apparent benefits, research in this area is still in its early stages. Most existing studies treat resource allocation and security separately, highlighting the need for further exploration of their integrated potential in semantic IoT systems. Future research should focus on developing innovative algorithms and mechanisms that seamlessly blend secure computation offloading, data sharing, and communication protocols while optimizing resource allocation. Additionally, analyzing the trade-off between resource allocation and security is essential to develop strategies that harmoniously balance these often conflicting objectives.

VII. Conclusion

In this comprehensive survey, we have ventured to delineate the resource management, security, and privacy in the context of SC. Our exploration has been grounded in a meticulous review of the existing literature, where we have highlighted both the strides made in the field and the open challenges that persist. Our contribution stands distinct in bringing together the discussions on resource management, security, and privacy under a single umbrella, while also offering a tutorial that elucidates the challenges, open problems, and potential future research directions. This endeavor marks a pioneering step in addressing these pivotal aspects collectively, aiming to foster a discourse that is rich and multifaceted. As we stand on the threshold of new developments in SC system, it becomes increasingly evident that further research and exploration are imperative to ensure the reliable and secure operation of these networks. The landscape of SC networks is ever-evolving, and it beckons a deeper dive into the mitigation strategies that can enhance their security and privacy.

Looking forward, we foresee a research trajectory that is vibrant and dynamic, encouraging scholars and practitioners alike to delve deeper and forge pathways that would steer SC networks towards a future that is not only efficient but also secure and reliable. The path forward offers numerous opportunities for innovation, inviting a new wave of researchers to develop solutions that are robust and sustainable. We conclude with a note of optimism, hopeful that this survey will serve as a catalyst for future research endeavors, fostering a landscape where SC networks can thrive, grounded in principles of efficiency, security, and privacy.

References

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

[2] C. Chaccour, M. N. Soorki, W. Saad, M. Bennis, P. Popovski, and M. Debbah, "Seven defining features of terahertz (thz) wireless systems: A fellowship of communication and sensing," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 967–993, 2022.

[3] E. C. Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.

[4] P. Zhang, W. Xu, H. Gao, K. Niu, X. Xu, X. Qin, C. Yuan, Z. Qin, H. Zhao, J. Wei *et al.*, "Toward wisdom-evolutionary and primitive-concise 6g: A new paradigm of semantic communication networks," *Engineering*, vol. 8, pp. 60–73, 2022.

[5] J. Liu, W. Zhang, and H. V. Poor, "A rate-distortion framework for characterizing semantic information," in *2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 2894–2899.

[6] W. Hong, C. Yu, J.-X. Chen, and Z. Hao, "Millimeter wave and terahertz technology," *Sci. China Inf. Sci*, vol. 46, pp. 1086–1107, 2016.

[7] W. Weaver, "Recent contributions to the mathematical theory of communication," *ETC: a review of general semantics*, pp. 261–281, 1953.

[8] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Resource allocation for text semantic communications," *IEEE Wireless Communications Letters*, vol. 11, no. 7, pp. 1394–1398, 2022.

[9] M. Kountouris and N. Pappas, "Semantics-empowered communication for networked intelligent systems," *IEEE Communications Magazine*, vol. 59, no. 6, pp. 96–102, 2021.

[10] 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, August 2021.

[11] C. Chaccour, W. Saad, M. Debbah, Z. Han, and H. V. Poor, "Less data, more knowledge: Building next generation semantic communication networks," *arXiv preprint arXiv:2211.14343*, 2022.

[12] Q. Lan, D. Wen, Z. Zhang, Q. Zeng, X. Chen, P. Popovski, and K. Huang, "What is semantic communication? a view on conveying meaning in the era of machine intelligence," *Journal of Communications and Information Networks*, vol. 6, no. 4, pp. 336–371, 2021.

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

[14] W. Yang, H. Du, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. S. Shen, and C. Miao, "Semantic communications for future internet: Fundamentals, applications, and challenges," *IEEE Communications Surveys & Tutorials*, 2022.

[15] T. M. Getu, G. Kaddoum, and M. Bennis, "Tutorial-cum-survey on semantic and goal- oriented communication: Research landscape, challenges, and future directions," 2023.

[16] Z. Yang, M. Chen, G. Li, Y. Yang, and Z. Zhang, "Secure semantic communications: Fundamentals and challenges," *arXiv preprint arXiv:2301.01421*, 2023.

[17] H. Du, J. Wang, D. Niyato, J. Kang, Z. Xiong, M. Guizani, and D. I. Kim, "Rethinking wireless communication security in semantic internet of things," *arXiv preprint arXiv:2210.04474*, 2022.

[18] E. C. Strinati and S. Barbarossa, "6g networks: Beyond shannon towards semantic and goal-oriented communications," *Computer Networks*, vol. 190, p. 107930, 2021.

[19] Z. Lu, R. Li, K. Lu, X. Chen, E. Hossain, Z. Zhao, and H. Zhang, "Semantics-empowered communications: A tutorial-cum-survey," *IEEE Communications Surveys & Tutorials*, 2023.

[20] N. Sharma and K. Kumar, "Resource allocation trends for ultra dense networks in 5g and beyond networks: A classification and comprehensive survey," *Physical Communication*, vol. 48, p. 101415, 2021.

[21] W. Ejaz, S. K. Sharma, S. Saadat, M. Naeem, A. Anpalagan, and N. A. Chughtai, "A comprehensive survey on resource allocation for cran in 5g and beyond networks," *Journal of Network and Computer Applications*, vol. 160, p. 102638, 2020.

[22] O. T. H. Alzubaidi, M. N. Hindia, K. Dimyati, K. A. Noordin, A. N. A. Wahab, F. Qamar, and R. Hassan, "Interference challenges and management in b5g network design: A comprehensive review," *Electronics*, vol. 11, no. 18, p. 2842, 2022.

[23] B. Agarwal, M. A. Togou, M. Marco, and G.-M. Muntean, "A comprehensive survey on radio resource management in 5g hetnets: Current solutions, future trends and open issues," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 4, pp. 2495–2534, 2022.

[24] S. Ebrahimi, F. Bouali, and O. C. Haas, "Resource management from single-domain 5g to end-to-end 6g network slicing: A survey," *IEEE Communications Surveys & Tutorials*, 2024.

[25] T. O. Olwal, K. Djouani, and A. M. Kurien, "A survey of resource management toward 5g radio access networks," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1656–1686, 2016.

[26] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, B. Soret, T. Soleymani, and K. H. Johansson, "Semantic communications in networked systems: A data significance perspective," *IEEE Network*, vol. 36, no. 4, pp. 233–240, 2022.

[27] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5g: Survey and challenges," *IEEE communications magazine*, vol. 55, no. 5, pp. 94–100, 2017.

[28] R. F. Olimid and G. Nencioni, "5g network slicing: A security overview," *IEEE Access*, vol. 8, pp. 99 999–100 009, 2020.

[29] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive mimo: An overview on passive eavesdropping and active attacks," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 21–27, 2015.

[30] M. Haus, M. Waqas, A. Y. Ding, Y. Li, S. Tarkoma, and J. Ott, "Security and privacy in device-to-device (d2d) communication: A review," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1054–1079, 2017.

[31] R. Khan, P. Kumar, D. N. K. Jayakody, and M. Liyanage, "A survey on security and privacy of 5g technologies: Potential solutions, recent advancements, and future directions," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 196–248, 2019.

[32] I. Ahmad, S. Shahabuddin, T. Kumar, J. Okwuibe, A. Gurtov, and M. Ylianttila, "Security for 5g and beyond," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3682–3722, 2019.

[33] I. Farris, T. Taleb, Y. Khettab, and J. Song, "A survey on emerging sdn and nfv security mechanisms for iot systems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 812–837, 2018.

[34] H.-M. Wang, T.-X. Zheng, J. Yuan, D. Towsley, and M. H. Lee, "Physical layer security in heterogeneous cellular networks," *IEEE Transactions on Communications*, vol. 64, no. 3, pp. 1204–1219, 2016.

[35] M. Schmittner, A. Asadi, and M. Hollick, "Semud: Secure multi-hop device-to-device communication for 5g public safety networks," in *2017 IFIP networking conference (IFIP Networking) and workshops*. IEEE, 2017, pp. 1–9.

[36] S. M. Alturfi, H. A. Marhoon, and B. Al-Musawi, "Internet of things security techniques: A survey," in *AIP Conference Proceedings*, vol. 2290, no. 1. AIP Publishing, 2020.

[37] F. I. Dretske, "Knowledge and the flow of information," 1981.

[38] J. Bao, P. Basu, M. Dean, C. Partridge, A. Swami, W. Leland, and J. A. Hendler, "Towards a theory of semantic communication," in *2011 IEEE Network Science Workshop*, 2011, pp. 110–117.

[39] J. Tang, Q. Yang, and Z. Zhang, "Information-theoretic limits on compression of semantic information," *arXiv preprint arXiv:2306.02305*, 2023.

[40] H. Rezaei, T. Sivalingam, and N. Rajatheva, "Automatic and flexible transmission of semantic map images using polar codes for end-to-end semantic-based communication systems," in *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2023, pp. 1–6.

[41] 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.

[42] ——, "Deep learning based semantic communications: An initial investigation," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

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

[44] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, and R. Q. Hu, "Cognitive semantic communication systems driven by knowledge graph," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 4860–4865.

[45] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications with masked vq-vae enabled codebook," *IEEE Transactions on Wireless Communications*, vol. 22, no. 12, pp. 8707–8722, 2023.

[46] H. Zhang, S. Shao, M. Tao, X. Bi, and K. B. Letaief, "Deep learning-enabled semantic communication systems with task-unaware transmitter and dynamic data," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 170–185, 2022.

[47] W. Zhang, H. Zhang, H. Ma, H. Shao, N. Wang, and V. C. M. Leung, "Predictive and adaptive deep coding for wireless image transmission in semantic communication," *IEEE Transactions on Wireless Communications*, vol. 22, no. 8, pp. 5486–5501, 2023.

[48] A. Varischio, F. Mandruzzato, M. Bullo, M. Giordani, P. Testolina, and M. Zorzi, "Hybrid point cloud semantic compression for automotive sensors: A performance evaluation," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[49] Y. Wang, P. H. Chan, and V. Donzella, "Semantic-aware video compression for automotive cameras," *IEEE Transactions on Intelligent Vehicles*, 2023.

[50] Z. Weng and Z. Qin, "Semantic communication systems for speech transmission," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2434–2444, 2021.

[51] H. Tong, Z. Yang, S. Wang, Y. Hu, O. Semiari, W. Saad, and C. Yin, "Federated learning for audio semantic communication," *Frontiers in communications and networks*, vol. 2, p. 734402, 2021.

[52] T. Han, Q. Yang, Z. Shi, S. He, and Z. Zhang, "Semantic-preserved communication system for highly efficient speech transmission," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 245–259, 2022.

[53] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2584–2597, 2022.

[54] A. Li, X. Wei, D. Wu, and L. Zhou, "Cross-modal semantic communications," *IEEE Wireless Communications*, vol. 29, no. 6, pp. 144–151, 2022.

[55] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, and X. Tao, "A unified multi-task semantic communication system for multimodal data," *IEEE Transactions on Communications*, 2024.

[56] N. Farsad, M. Rao, and A. Goldsmith, "Deep learning for joint source-channel coding of text," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 2326–2330.

[57] 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.

[58] D. B. Kurka and D. Gündüz, "Deepjscc-f: Deep joint source-channel coding of images with feedback," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 1, pp. 178–193, 2020.

[59] J. Xu, B. Ai, W. Chen, A. Yang, P. Sun, and M. Rodrigues, "Wireless image transmission using deep source channel coding with attention modules," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2315–2328, 2021.

[60] Google, "Draco 3D Data Compression," https://github.com/google/draco, 2017.

[61] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet++: Fast and accurate lidar semantic segmentation," in *2019 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 2019, pp. 4213–4220.

[62] M. Chen, M. Liu, W. Wang, H. Dou, and L. Wang, "Cross-modal semantic communications in 6g," in *2023 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, 2023, pp. 1–6.

[63] H. Xie, Z. Qin, and G. Y. Li, "Task-oriented multi-user semantic communications for vqa," *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 553–557, 2021.

[64] M. Rao, N. Farsad, and A. Goldsmith, "Variable length joint source-channel coding of text using deep neural networks," in *2018 IEEE 19th international workshop on signal processing advances in wireless communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[65] M. Sana and E. C. Strinati, "Learning semantics: An opportunity for effective 6g communications," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, 2022, pp. 631–636.

[66] P. Jiang, C.-K. Wen, S. Jin, and G. Y. Li, "Deep source-channel coding for sentence semantic transmission with harq," *IEEE transactions on communications*, vol. 70, no. 8, pp. 5225–5240, 2022.

[67] Q. Zhou, R. Li, Z. Zhao, Y. Xiao, and H. Zhang, "Adaptive bit rate control in semantic communication with incremental knowledge-based harq," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1076–1089, 2022.

[68] Y. Zhu, Y. Huang, X. Qiao, Z. Tan, B. Bai, H. Ma, and S. Dustdar, "A semantic-aware transmission with adaptive control scheme for volumetric video service," *IEEE Transactions on Multimedia*, pp. 1–13, 2022.

[69] K. Chi, Q. Yang, Z. Yang, Y. Duan, and Z. Zhang, "Resource allocation for capacity optimization in joint source-channel coding systems," 2022.

[70] C. Liu, C. Guo, Y. Yang, and N. Jiang, "Adaptable semantic compression and resource allocation for task-oriented communications," 2022.

[71] F. Binucci, P. Banelli, P. Di Lorenzo, and S. Barbarossa, "Adaptive resource optimization for edge inference with goal-oriented communications," *EURASIP Journal on Advances in Signal Processing*, vol. 2022, no. 1, p. 123, 2022.

[72] F. Binucci, P. Banelli, P. D. Lorenzo, and S. Barbarossa, "Multi-user goal-oriented communications with energy-efficient edge resource management," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2023.

[73] L. Yan, Z. Qin, R. Zhang, Y. Li, and G. Y. Li, "Qoe-aware resource allocation for semantic communication networks," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 3272–3277.

[74] Z. Yang, M. Chen, Z. Zhang, and C. Huang, "Energy efficient semantic communication over wireless networks with rate splitting," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 5, pp. 1484–1495, 2023.

[75] W. Li, H. Liang, C. Dong, X. Xu, P. Zhang, and K. Liu, "Non-orthogonal multiple access enhanced multi-user semantic communication," 2023.

[76] H. Zhang, H. Wang, Y. Li, K. Long, and A. Nallanathan, "Drl-driven dynamic resource allocation for task-oriented semantic communication," *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 3992–4004, July 2023.

[77] L. Xia, Y. Sun, X. Li, G. Feng, and M. A. Imran, "Wireless resource management in intelligent semantic communication networks," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2022, pp. 1–6.

[78] X. Mu and Y. Liu, "Semi-noma enabled coexisting semantic and bit communications," in *2022 International Symposium on Wireless Communication Systems (ISWCS)*. IEEE, 2022, pp. 1–6.

[79] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2584–2597, Sept. 2022.

[80] X. Mu and Y. Liu, "Exploiting semantic communication for non-orthogonal multiple access," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 8, pp. 2563–2576, 2023.

[81] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[82] J. Kang, H. Du, Z. Li, Z. Xiong, S. Ma, D. Niyato, and Y. Li, "Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 186–201, 2022.

[83] B. Du, H. Du, H. Liu, D. Niyato, P. Xin, J. Yu, M. Qi, and Y. Tang, "Yolo-based semantic communication with generative ai-aided resource allocation for digital twins construction," *arXiv preprint arXiv:2306.14138*, 2023.

[84] P. Jiang, C.-K. Wen, S. Jin, and G. Li, "Wireless semantic communications for video conferencing," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 230–244, Jan. 2023.

[85] M. K. Farshbafan, W. Saad, and M. Debbah, "Curriculum learning for goal-oriented semantic communications with a common language," *IEEE Transactions on Communications*, vol. 71, no. 3, pp. 1430–1446, March 2023.

[86] Y. Wang, M. Chen, T. Luo, W. Saad, D. Niyato, H. V. Poor, and S. Cui, "Performance optimization for semantic communications: An attention-based reinforcement learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2598–2613, 2022.

[87] H. Zhang, H. Wang, Y. Li, K. Long, and V. C. M. Leung, "Toward intelligent resource allocation on task-oriented semantic communication," *IEEE Wireless Communications*, vol. 30, no. 3, pp. 70–77, June 2023.

[88] S. Wang, S. Bi, and Y.-J. A. Zhang, "Deep reinforcement learning with communication transformer for adaptive live streaming in wireless edge networks," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 308–322, Jan. 2022.

[89] H. Du, J. Liu, D. Niyato, J. Kang, Z. Xiong, J. Zhang, and D. I. Kim, "Attention-aware resource allocation and qoe analysis for metaverse xurllc services," *IEEE Journal on Selected Areas in Communications*, 2023.

[90] J. Joshua, "Information bodies: computational anxiety in neal stephenson's snow crash," *Interdisciplinary Literary Studies*, vol. 19, no. 1, pp. 17–47, 2017.

[91] Z. Ji and Z. Qin, "Energy-efficient task offloading for semantic-aware networks," *arXiv preprint arXiv:2301.08376*, 2023.

[92] W. Yang, Z. Q. Liew, W. Y. B. Lim, Z. Xiong, D. Niyato, X. Chi, X. Cao, and K. B. Letaief, "Semantic communication meets edge intelligence," *IEEE Wireless Communications*, vol. 29, no. 5, pp. 28–35, 2022.

[93] X. Wang, Y. Han, V. C. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 869–904, 2020.

[94] R. Fantacci and B. Picano, "Multi-user semantic communications system with spectrum scarcity," *Journal of Communications and Information Networks*, vol. 7, no. 4, pp. 375–382, Dec. 2022.

[95] R. Khan, D. N. K. Jayakody, H. Pervaiz, and R. Tafazolli, "Modulation based non-orthogonal multiple access for 5g resilient networks," in *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2018, pp. 1–6.

[96] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, "Heterogeneous semantic and bit communications: A semi-noma scheme," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 155–169, 2023.

[97] X. Mu and Y. Liu, "Semantic communications in multi-user wireless networks," *arXiv preprint arXiv:2211.08932*, 2022.

[98] J. Chen, J. Wang, C. Jiang, and J. Wang, "Age of incorrect information in semantic communications for noma aided xr applications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 5, pp. 1093–1105, 2023.

[99] L. Yan, Z. Qin, R. Zhang, Y. Li, X. Tao, and G. Y. Li, "Qoe-based semantic-aware resource allocation for multi-task networks," *arXiv preprint arXiv:2305.06543*, 2023.

[100] A. Ndikumana, N. H. Tran, T. M. Ho, Z. Han, W. Saad, D. Niyato, and C. S. Hong, "Joint communication, computation, caching, and control in big data multi-access edge computing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 6, pp. 1359–1374, 2019.

[101] Z. Ji, Z. Qin, X. Tao, and Z. Han, "Resource optimization for semantic-aware networks with task offloading," *IEEE Transactions on Wireless Communications*, 2024.

[102] C. Wang *et al.*, "Multimodal semantic communication accelerated bidirectional caching for 6g mec," *Future Generation Computer Systems*, vol. 140, pp. 225–237, 2023.

[103] Y. Che, H. Xiong, S. Han, and X. Xu, "Cache-enabled knowledge base construction strategy in semantic communications," in *2022 IEEE Globecom Workshops (GC Wkshps)*, 2022, pp. 1507–1512.

[104] P. Popovski, O. Simeone, F. Boccardi, D. Gündüz, and O. Sahin, "Semantic-effectiveness filtering and control for post-5g wireless connectivity," *Journal of the Indian Institute of Science*, vol. 100, 05 2020.

[105] T. N. Dang, A. Manzoor, Y. K. Tun, S. M. A. Kazmi, R. Haw, S. H. Hong, Z. Han, and C. S. Hong, "Joint communication, computation, and control for computational task offloading in vehicle-assisted multi-access edge computing," *IEEE Access*, vol. 10, pp. 122 513–122 529, 2022.

[106] Z. A. Hmitti, H. B. Ammar, E. G. Soyak, Y. Kardjadja, S. Malektaji, S. O. Ali, M. Rayani, M. Saqib, S. Taghizadeh, W. Ajib, H. Elbiaze, O. Ercetin, Y. Ghamri-Doudane, and R. Glitho, "Scoring: Towards smart collaborative computing, caching and networking paradigm for next generation communication infrastructures," in *2022 International Conference on Computer Communications and Networks (ICCCN)*, 2022, pp. 1–10.

[107] M. Xue, C. Yuan, H. Wu, Y. Zhang, and W. Liu, "Machine learning security: Threats, countermeasures, and evaluations," *IEEE Access*, vol. 8, pp. 74 720–74 742, 2020.

[108] M. Goldblum, D. Tsipras, C. Xie, X. Chen, A. Schwarzschild, D. Song, A. Madry, B. Li, and T. Goldstein, "Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1563–1580, 2023.

[109] M. Lokumarambage, V. Gowrisetty, H. Rezaei, T. Sivalingam, N. Rajatheva, and A. Fernando, "Wireless end-to-end image transmission system using semantic communications," *IEEE Access*, 2023.

[110] 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.

[111] J. Chen, L. Zhang, H. Zheng, X. Wang, and Z. Ming, "Deeppoison: Feature transfer based stealthy poisoning attack for dnns," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 7, pp. 2618–2622, 2021.

[112] S. Xie, Y. Yan, and Y. Hong, "Stealthy 3d poisoning attack on video recognition models," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[113] Z. Weng, Z. Qin, X. Tao, C. Pan, G. Liu, and G. Y. Li, "Deep learning enabled semantic communications with speech recognition and synthesis," *IEEE Transactions on Wireless Communications*, 2023.

[114] H. Aghakhani, T. Eisenhofer, L. Schönherr, D. Kolossa, T. Holz, C. Kruegel, and G. Vigna, "Venomave: Clean-label poisoning against speech recognition," *Computing Research Repository (CoRR), abs/2010.10682*, 2020.

[115] A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha, "Secure and robust machine learning for healthcare: A survey," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 156–180, 2020.

[116] Y. E. Sagduyu, T. Erpek, S. Ulukus, and A. Yener, "Vulnerabilities of deep learning-driven semantic communications to backdoor (trojan) attacks," in *2023 57th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2023, pp. 1–6.

[117] J. Zhang, C. Dongdong, Q. Huang, J. Liao, W. Zhang, H. Feng, G. Hua, and N. Yu, "Poison ink: Robust and invisible backdoor attack," *IEEE Transactions on Image Processing*, vol. 31, pp. 5691–5705, 2022.

[118] J. Jia, Y. Liu, and N. Z. Gong, "Badencoder: Backdoor attacks to pre-trained encoders in self-supervised learning," in *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2022, pp. 2043–2059.

[119] R. Wang, H. Chen, Z. Zhu, L. Liu, Y. Zhang, Y. Fan, and B. Wu, "Robust backdoor attack with visible, semantic, sample-specific, and compatible triggers," *arXiv preprint arXiv:2306.00816*, 2023.

[120] J. Dai, C. Chen, and Y. Li, "A backdoor attack against lstm-based text classification systems," *IEEE Access*, vol. 7, pp. 138 872–138 878, 2019.

[121] J. Yan, V. Gupta, and X. Ren, "Bite: Textual backdoor attacks with iterative trigger injection," in *ICLR 2023 Workshop on Backdoor Attacks and Defenses in Machine Learning*, 2023.

[122] X. Liu, L. Xie, Y. Wang, J. Zou, J. Xiong, Z. Ying, and A. V. Vasilakos, "Privacy and security issues in deep learning: A survey," *IEEE Access*, vol. 9, pp. 4566–4593, 2020.

[123] A. Rahman, M. S. Hossain, N. A. Alrajeh, and F. Alsolami, "Adversarial examples—security threats to covid-19 deep learning systems in medical iot devices," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9603–9610, 2020.

[124] A. Bar, J. Lohdefink, N. Kapoor, S. J. Varghese, F. Huger, P. Schlicht, and T. Fingscheidt, "The vulnerability of semantic segmentation networks to adversarial attacks in autonomous driving: Enhancing extensive environment sensing," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 42–52, 2020.

[125] H. Wu, S. Yunas, S. Rowlands, W. Ruan, and J. Wahlstrom, "Adversarial driving: Attacking end-to-end autonomous driving," *arXiv preprint arXiv:2103.09151*, 2021.

[126] S. Y. Khamaiseh, D. Bagagem, A. Al-Alaj, M. Mancino, and H. W. Alomari, "Adversarial deep learning: A survey on adversarial attacks and defense mechanisms on image classification," *IEEE Access*, 2022.

[127] X. Yang, W. Liu, S. Zhang, W. Liu, and D. Tao, "Targeted attention attack on deep learning models in road sign recognition," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4980–4990, 2020.

[128] Q. Wang, B. Zheng, Q. Li, C. Shen, and Z. Ba, "Towards query-efficient adversarial attacks against automatic speech recognition systems," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 896–908, 2020.

[129] S. Wang, Z. Zhang, G. Zhu, X. Zhang, Y. Zhou, and J. Huang, "Query-efficient adversarial attack with low perturbation against end-to-end speech recognition systems," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 351–364, 2022.

[130] M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart, "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, 2014, pp. 17–32.

[131] Y. Zhang, R. Jia, H. Pei, W. Wang, B. Li, and D. Song, "The secret revealer: Generative model-inversion attacks against deep neural networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 253–261.

[132] K. Mo, T. Huang, and X. Xiang, "Querying little is enough: Model inversion attack via latent information," in *Machine Learning for Cyber Security: Third International Conference, ML4CS 2020, Guangzhou, China, October 8–10, 2020, Proceedings, Part II 3*. Springer, 2020, pp. 583–591.

[133] R. Zhang, S. Hidano, and F. Koushanfar, "Text revealer: Private text reconstruction via model inversion attacks against transformers," *arXiv preprint arXiv:2209.10505*, 2022.

[134] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang, "Membership inference attacks on machine learning: A survey," *ACM Computing Surveys (CSUR)*, vol. 54, no. 11s, pp. 1–37, 2022.

[135] A. Jagannatha, B. P. S. Rawat, and H. Yu, "Membership inference attack susceptibility of clinical language models," 2021.

[136] M. Zhang, Z. Ren, Z. Wang, P. Ren, Z. Chen, P. Hu, and Y. Zhang, "Membership inference attacks against recommender systems," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 864–879.

[137] X. Peng, Z. Qin, D. Huang, X. Tao, J. Lu, G. Liu, and C. Pan, "A robust deep learning enabled semantic communication system for text," in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 2704–2709.

[138] Q. Hu, G. Zhang, Z. Qin, Y. Cai, G. Yu, and G. Y. Li, "Robust semantic communications with masked vq-vae enabled codebook," *IEEE Transactions on Wireless Communications*, 2023.

[139] ——, "Robust semantic communications against semantic noise," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. IEEE, 2022, pp. 1–6.

[140] X. Wei, Y. Guo, and J. Yu, "Adversarial sticker: A stealthy attack method in the physical world," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[141] A. Li, X. Wei, D. Wu, and L. Zhou, "Cross-modal semantic communications," *IEEE Wireless Communications*, vol. 29, no. 6, pp. 144–151, 2022.

[142] J. Choi and J. Park, "Semantic communication as a signaling game with correlated knowledge bases," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*. IEEE, 2022, pp. 1–5.

[143] S. Ji, S. Pan, E. Cambria, P. Marttinen, and S. Y. Philip, "A survey on knowledge graphs: Representation, acquisition, and applications," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 2, pp. 494–514, 2021.

[144] F. Zhou, Y. Li, X. Zhang, Q. Wu, X. Lei, and R. Q. Hu, "Cognitive semantic communication systems driven by knowledge graph," in *ICC 2022-IEEE International Conference on Communications*. IEEE, 2022, pp. 4860–4865.

[145] Y. Wang, Z. Su, S. Guo, M. Dai, T. H. Luan, and Y. Liu, "A survey on digital twins: architecture, enabling technologies, security and privacy, and future prospects," *IEEE Internet of Things Journal*, 2023.

[146] D. Wheeler and B. Natarajan, "Engineering semantic communication: A survey," *IEEE Access*, vol. 11, pp. 13 965–13 995, 2023.

[147] Z.-W. Wu, C.-T. Chen, and S.-H. Huang, "Poisoning attacks against knowledge graph-based recommendation systems using deep reinforcement learning," *Neural Computing and Applications*, pp. 1–19, 2022.

[148] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2584–2597, 2022.

[149] A. R. Shahid, A. Imteaj, P. Y. Wu, D. A. Igoche, and T. Alam, "Label flipping data poisoning attack against wearable human activity recognition system," in *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2022, pp. 908–914.

[150] Y. Ge, Q. Wang, J. Yu, C. Shen, and Q. Li, "Data poisoning and backdoor attacks on audio intelligence systems," *IEEE Communications Magazine*, pp. 1–7, 2023.

[151] Y. Lin, Z. Gao, H. Du, D. Niyato, J. Kang, A. Jamalipour, and X. S. Shen, "A unified framework for integrating semantic communication and ai-generated content in metaverse," *arXiv preprint arXiv:2305.11911*, 2023.

[152] M. Ali, F. Naeem, G. Kaddoum, and E. Hossain, "Metaverse communications, networking, security, and applications: Research issues, state-of-the-art, and future directions," *arXiv preprint arXiv:2212.13993*, 2022.

[153] E. Uysal, O. Kaya, A. Ephremides, J. Gross, M. Codreanu, P. Popovski, M. Assaad, G. Liva, A. Munari, B. Soret *et al.*, "Semantic communications in networked systems: A data significance perspective," *IEEE Network*, vol. 36, no. 4, pp. 233–240, 2022.

[154] Z. Zhang, R. Deng, D. K. Y. Yau, and P. Chen, "Zero-parameter-information data integrity attacks and countermeasures in iot-based smart grid," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6608–6623, 2021.

[155] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

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

[157] X. Luo, H.-H. Chen, and Q. Guo, "Semantic communications: Overview, open issues, and future research directions," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 210–219, 2022.

[158] G. Nan, Z. Li, J. Zhai, Q. Cui, G. Chen, X. Du, X. Zhang, X. Tao, Z. Han, and T. Q. Quek, "Physical-layer adversarial robustness for deep learning-based semantic communications," *IEEE Journal on Selected Areas in Communications*, 2023.

[159] M. Sadeghi and E. G. Larsson, "Physical adversarial attacks against end-to-end autoencoder communication systems," *IEEE Communications Letters*, vol. 23, no. 5, pp. 847–850, 2019.

[160] 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.

[161] J. Kang, J. He, H. Du, Z. Xiong, Z. Yang, X. Huang, and S. Xie, "Adversarial attacks and defenses for semantic communication in vehicular metaverses," *arXiv preprint arXiv:2306.03528*, 2023.

[162] O. Salem, K. Alsubhi, A. Shaafi, M. Gheryani, A. Mehaoua, and R. Boutaba, "Man-in-the-middle attack mitigation in internet of medical things," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2053–2062, 2022.

[163] P. Solnør, Ø. Volden, K. Gryte, S. Petrovic, and T. I. Fossen, "Hijacking of unmanned surface vehicles: A demonstration of attacks and countermeasures in the field," *Journal of Field Robotics*, vol. 39, no. 5, pp. 631–649, 2022.

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

[165] R. Tang, D. Gao, M. Yang, T. Guo, H. Wu, and G. Shi, "Gan-inspired intelligent jamming and anti-jamming strategy for semantic communication systems," 05 2023.

[166] J. Chen, X. Zhang, R. Zhang, C. Wang, and L. Liu, "De-pois: An attack-agnostic defense against data poisoning attacks," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3412–3425, 2021.

[167] J. Geiping, L. Fowl, G. Somepalli, M. Goldblum, M. Moeller, and T. Goldstein, "What doesn't kill you makes you robust (er): How to adversarially train against data poisoning," *arXiv preprint arXiv:2102.13624*, 2021.

[168] H. Huang, W. Luo, G. Zeng, J. Weng, Y. Zhang, and A. Yang, "Damia: Leveraging domain adaptation as a defense against membership inference attacks," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 5, pp. 3183–3199, 2022.

[169] R. Theagarajan and B. Bhanu, "Privacy preserving defense for black box classifiers against on-line adversarial attacks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 9503–9520, 2021.

[170] G. Chen, Z. Zhao, F. Song, S. Chen, L. Fan, F. Wang, and J. Wang, "Towards understanding and mitigating audio adversarial examples for speaker recognition," *IEEE Transactions on Dependable and Secure Computing*, 2022.

[171] I.-H. Liu, J.-S. Li, Y.-C. Peng, and C.-G. Liu, "A robust countermeasures for poisoning attacks on deep neural networks of computer interaction systems," *Applied Sciences*, vol. 12, no. 15, p. 7753, 2022.

[172] 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.

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

[174] X. Luo, Z. Chen, M. Tao, and F. Yang, "Encrypted semantic communication using adversarial training for privacy preserving," *IEEE Communications Letters*, vol. 27, no. 6, pp. 1486–1490, 2023.

[175] 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.

[176] Q. Qin, Y. Rong, G. Nan, S. Wu, X. Zhang, Q. Cui, and X. Tao, "Securing semantic communications with physical-layer semantic encryption and obfuscation," *arXiv preprint arXiv:2304.10147*, 2023.

[177] A. Pourranjbar, G. Kaddoum, and W. Saad, "Recurrent neural network-based anti-jamming framework for defense against multiple jamming policies," *IEEE Internet of Things Journal*, 2023.

[178] A. Deb Raha, M. Shirajum Munir, A. Adhikary, Y. Qiao, S.-B. Park, and C. Seon Hong, "An artificial intelligent-driven semantic communication framework for connected autonomous vehicular network," in *2023 International Conference on Information Networking (ICOIN)*, 2023, pp. 352–357.

[179] Y. Yao, J. Zhao, Z. Li, X. Cheng, and L. Wu, "Jamming and eavesdropping defense scheme based on deep reinforcement learning in autonomous vehicle networks," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 1211–1224, 2023.

[180] P. R. Babu, R. Amin, A. G. Reddy, A. K. Das, W. Susilo, and Y. Park, "Robust authentication protocol for dynamic charging system of electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 11, pp. 11 338–11 351, 2021.

[181] W. Xu, Y. Zhang, F. Wang, Z. Qin, C. Liu, and P. Zhang, "Semantic communication for the internet of vehicles: A multiuser cooperative approach," *IEEE Vehicular Technology Magazine*, vol. 18, no. 1, pp. 100–109, 2023.

[182] H. Kwon, "Defending deep neural networks against backdoor attack by using de-trigger autoencoder," *IEEE Access*, 2021.

[183] Z. Xiang, D. J. Miller, and G. Kesidis, "Detection of backdoors in trained classifiers without access to the training set," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 3, pp. 1177–1191, 2020.

[184] S. Wang, S. Nepal, C. Rudolph, M. Grobler, S. Chen, T. Chen, and Z. An, "Defending adversarial attacks via semantic feature manipulation," *IEEE Transactions on Services Computing*, vol. 15, no. 6, pp. 3184–3197, 2021.

[185] Y. Alufaisan, M. Kantarcioglu, and Y. Zhou, "Robust transparency against model inversion attacks," *IEEE transactions on dependable and secure computing*, vol. 18, no. 5, pp. 2061–2073, 2020.

[186] D. Ye, S. Shen, T. Zhu, B. Liu, and W. Zhou, "One parameter defense – defending against data inference attacks via differential privacy," 2022.

[187] C. Sey, H. Lei, W. Qian, X. Li, L. D. Fiasam, S. L. Kodjiku, I. Adjei-Mensah, and I. O. Agyemang, "Vblock: A blockchain-based tamper-proofing data protection model for internet of vehicle networks," *Sensors*, vol. 22, no. 20, p. 8083, 2022.

[188] R. Kaewpuang, M. Xu, W. Y. B. Lim, D. Niyato, H. Yu, J. Kang, and X. S. Shen, "Cooperative resource management in quantum key distribution (qkd) networks for semantic communication," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[189] Y. Cao, Y. Zhao, Q. Wang, J. Zhang, S. X. Ng, and L. Hanzo, "The evolution of quantum key distribution networks: On the road to the

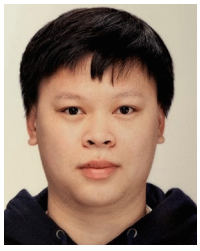
qinternet," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 839–894, 2022.
[190] A. S. Cacciapuoti, M. Caleffi, F. Tafuri, F. S. Cataliotti, S. Gherardini, and G. Bianchi, "Quantum internet: Networking challenges in distributed quantum computing," *IEEE Network*, vol. 34, no. 1, pp. 137–143, 2019.
[191] A. S. Cacciapuoti, M. Caleffi, R. Van Meter, and L. Hanzo, "When entanglement meets classical communications: Quantum teleportation for the quantum internet," *IEEE Transactions on Communications*, vol. 68, no. 6, pp. 3808–3833, 2020.
[192] J. Illiano, M. Caleffi, A. Manzalini, and A. S. Cacciapuoti, "Quantum internet protocol stack: A comprehensive survey," *Computer Networks*, vol. 213, p. 109092, 2022.
[193] C. Dong, H. Liang, X. Xu, S. Han, B. Wang, and P. Zhang, "Innovative semantic communication system," *arXiv preprint arXiv:2202.09595*, 2022.
[194] J. Kang, H. Du, Z. Li, Z. Xiong, S. Ma, D. Niyato, and Y. Li, "Personalized saliency in task-oriented semantic communications: Image transmission and performance analysis," *IEEE Journal on Selected Areas in Communications*, vol. 41, no. 1, pp. 186–201, 2022.
[195] Q. Sun, C. Guo, Y. Yang, J. Chen, and X. Xue, "Semantic-assisted image compression," *arXiv preprint arXiv:2201.12599*, 2022.
[196] L. X. Nguyen, K. Kim, Y. L. Tun, S. S. Hassan, Y. K. Tun, Z. Han, and C. S. Hong, "Optimizing multi-user semantic communication via transfer learning and knowledge distillation," *arXiv preprint arXiv:2406.03773*, 2024.
[197] Q. Wu, F. Liu, H. Xia, and T. Zhang, "Semantic transfer between different tasks in the semantic communication system," in *2022 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2022, pp. 566–571.
[198] M. Maier, A. Ebrahimzadeh, S. Rostami, and A. Beniiche, "The internet of no things: Making the internet disappear and" see the invisible"," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 76–82, 2020.
[199] P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations and trends in multimodal machine learning: Principles, challenges, and open questions," *arXiv preprint arXiv:2209.03430*, 2022.
[200] P. Wang, J. Li, C. Liu, X. Fan, M. Ma, and Y. Wang, "Distributed semantic communications for multimodal audio-visual parsing tasks," *IEEE Transactions on Green Communications and Networking*, 2024.
[201] Z. Qin, G. Y. Li, and H. Ye, "Federated learning and wireless communications," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 134–140, 2021.
[202] G. Zheng, Q. Ni, K. Navaie, and H. Pervaiz, "Semantic communication in satellite-borne edge cloud network for computation offloading," *IEEE Journal on Selected Areas in Communications*, 2024.
[203] H. Xie, Z. Qin, X. Tao, and K. B. Letaief, "Task-oriented multi-user semantic communications," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2584–2597, 2022.
[204] A. Zhang, Y. Wang, and S. Guo, "On the utility-informativeness-security trade-off in discrete task-oriented semantic communication," *IEEE Communications Letters*, 2024.
[205] J. Shao, Y. Mao, and J. Zhang, "Task-oriented communication for multidevice cooperative edge inference," *IEEE Transactions on Wireless Communications*, vol. 22, no. 1, pp. 73–87, 2022.

Dongwook Won received a B.S. and M.S. degree in Electronics and Communications Engineering from Kwangwoon University, Seoul, Korea in 2020, 2022. He is currently pursuing a Ph.D. in the School of Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include Semantic Communications, 5G/6G Networks, and Satellite Communications.

Geeranuch Woraphonbenjakul received the B.S. degree in telecommunications engineering from King Mongkut's University of Technology North Bangkok, Bangkok, Thailand, in 2016. She then obtained the M.S. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2023. Her research interests include wireless sensor networks and wireless communication.

Ayalneh Bitew Wondmagegn received the B.S. degree in Information Technology from Addis Ababa University, Addis Ababa, Ethiopia in 2005, the M.S. degree in Information Management from Hradec Kralove University, Hradec Kralove, Czech Republic in 2012, He is currently pursuing the Ph.D. degree in computer science and engineering with the School of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea. He was a Lecturer with the College of Technology, Debre Markos University, Debre Markos, Ethiopia. His research interests include wireless communication, Internet of Things, mobile edge computing, reinforcement learning, and flying ad hoc networks.

Donghyun Lee received his B.S. degree in Computer Science and Engineering from Dongguk University, South Korea, in 2020. He received M.S. degrees in Computer Science and Engineering from Chung-Ang University, Seoul, South Korea, in 2022. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include wireless networks, multimedia communications, and multiple access.



Anh-Tien Tran received the B.S. degree in electronics and telecommunications from the Da Nang University of Science and Technology, Da Nang, Vietnam, in 2018. He is currently working toward the Ph.D. degree in computer science and engineering with Chung-Ang University, Seoul, South Korea. His research interests include wireless next generation network communication, video streaming, and machine learning.

Demeke Shumeye Lakew received the B.S. degree in Computer Science from Hawassa University, Hawassa, Ethiopia in 2006, the M.S. degree in Computer Science from Addis Ababa University, Addis Ababa, Ethiopia in 2011, and the Ph.D. degree in Computer Science and Engineering from the School of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea, in 2023. He is currently an Assistant Professor with the Department of Computer Science, Kombolcha Institute of Technology, Wollo University, Dessie, Ethiopia. His research interests include wireless communication, mobile edge computing, reinforcement learning, Internet of Things, and flying ad hoc networks.

Sungrae Cho received the B.S. and M.S. degrees in Electronics Engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002. He is a Professor with the School of Computer Sciences and Engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an Assistant Professor with the Department of Computer Sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a Senior Member of Technical Staff with the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a Research Staff Member with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a Visiting Professorship with the National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA. He has been a KICS fellow since 2021. He received numerous awards including Haedong Best Researcher of 2022 in Telecommunications and Award of Korean Ministry of Science and ICT in 2021. His current research interests include wireless networking, network intelligence, and network optimization. He has been an Editor-in-Chief (EIC) of ICT Express (Elsevier) since 2024, a Subject Editor of IET Electronics Letter since 2018, an Executive Editor of Wiley Transactions on Emerging Telecommunications Technologies since 2023, and was an Area Editor of Ad Hoc Networks Journal (Elsevier) from 2012 to 2017. He has served numerous international conferences as a general chair, TPC chair, or an organizing committee chair, such as IEEE ICC, IEEE SECON, IEEE ICCE, ICOIN, ICTC, ICUFN, APCC, TridentCom, and the IEEE MASS.