Toward Mixture-of-Experts Enabled Trustworthy Semantic Communication for 6G Networks

Jiayi He, Xiaofeng Luo, Jiawen Kang, Hongyang Du, Zehui Xiong, Ci Chen, Dusit Niyato, and Xuemin Shen

Abstract

Semantic Communication (SemCom) plays a pivotal role in 6G networks, offering a viable solution for future efficient communication. Through nonlinear mapping of semantic representations, Deep Learning (DL)-based semantic communications further enhance communication efficiency. However, the vulnerability of DL models to security threats, such as adversarial attacks, poses significant challenges for practical applications of SemCom systems. These vulnerabilities allow attackers to tamper with messages and eavesdrop on private information, especially in wireless communication scenarios. Although existing defenses attempt to address specific threats, they often fail to simultaneously handle multiple heterogeneous attacks. To overcome this limitation, we propose a novel Mixture-of-Experts (MoE)-based SemCom system, comprising a gating network and multiple experts, each specializing in different security challenges. The gating network adaptively selects suitable experts to counter heterogeneous attacks based on user-defined security requirements. Multiple experts collaborate to accomplish semantic communication tasks while meeting the security requirements of users. A case study in vehicular networks demonstrates the efficacy of the MoE-based SemCom system. Simulation results show that the proposed MoE-based SemCom system effectively mitigates concurrent heterogeneous attacks, with minimal impact on downstream task accuracy.

Introduction

In the era of 6G, the integration of networks spanning space, air, ground, and sea is anticipated to facilitate instant, efficient, and intelligent hyper-connectivity communications. Semantic Communication (SemCom) emerges as a crucial component in 6G, serving as an efficient and intelligent paradigm for future communication [1]. The adoption of an encoder-decoder structure within the SemCom system allows for the transmission of only essential information in source messages, enabling the receiver to accurately reconstruct the original message. With this structure, SemCom surpasses the Shannon Limit, accomplishing communication tasks with minimal bit transmission. Consequently, SemCom is poised to bolster end-to-end wireless communications in scenarios involving a high volume of immersive communication tasks, such as metaverses, in environments with stringent delay requirements like autonomous driving, and in situations where bandwidth resources are limited, such as maritime communication and remote area communication, especially in the 6G landscape [1], [2], [3].

The evolution of Deep Learning (DL) models has significantly improved the performance of SemCom systems, enabling them to handle a variety of complex tasks, such as multimodal [4] and task-oriented communications [5]. However, this progress also highlights inherent vulnerabilities in SemCom systems, primarily linked to the weaknesses of DL models. One critical vulnerability is their susceptibility to semantic noise attacks, where an attacker injects semantic noise into the communication source, leading to decoded message distortion or degraded accuracy [6]. The open nature of wireless channels exacerbates this fragility, as it permits the injection of semantic noise into the physical channel [7], [8]. Moreover, the presence of a warden in the communication environment poses additional privacy risks. A warden can monitor active communications between devices and exploit exposed model parameters of the semantic decoder to extract sensitive user information [9], [10]. These weaknesses can easily be exploited to tamper with critical information or eavesdrop on private information. For instance, in a SemCom-based vehicular network, a well-behaved vehicle broadcasts safety messages to nearby vehicles to enhance driving safety [11]. However, if an attacker tampers with the message by adding semantic noise, surrounding vehicles will receive falsified information. This may cause unnecessary emergency braking, leading vehicles into dangerous situations. Therefore, investigating trustworthy semantic communication in 6G networks is urgently needed.

There are several approaches to address the aforementioned vulnerabilities. To tackle semantic noise attacks, Kang et al. [6] and Nan et al. [7] suggest enhancing the robustness of semantic codecs by using adversarial training technology.

Jiayi He, Xiaofeng Luo, Jiawen Kang (corresponding author), and Ci Chen are with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China; Hongyang Du and Dusit Niyato are with the College of Computing and Data Science, Nanyang Technological University, Singapore 639798; Zehui Xiong is with the Pillar of Information Systems Technology and Design, Singapore University of Technology and Design, Singapore 487372; Xuemin Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada. Jiayi He and Xiaofeng Luo contributed equally to this work.

Digital Object Identifier: 10.1109/MNET.2024.3523181
Date of Current Version: 16 September 2025
Date of Publication: 26 December 2024

SemCom concentrates on extracting the semantic contents of the source message at the transmitter and interpreting their meanings at the receiver.

They incorporate adversarial semantics noise into the source or channel during the training of the semantic codec, which can bolster the system's resistance against semantic noise attacks originating from both the source and the physical channel. To shield communication status from detection by a purposeful warden, Du et al. [9] propose that the semantic compression rate can be adjusted to reduce the probability of communication behavior being detected. Furthermore, to safeguard against illegal decoding of the semantic information, Luo et al. [10] advocate deploying a semantic encryption mechanism. Despite these advancements, the challenge of developing a SemCom system that can effectively counteract multiple security threats simultaneously remains largely unexplored in the evolution toward trustworthy 6G networks.

The Mixture-of-Experts (MoE) model presents a compelling approach to enhance trustworthy SemCom, particularly in the face of multiple heterogeneous security challenges. By partitioning the model and dataset into multiple specialized subgroups, the MoE framework accommodates tailored processing of diverse inputs [12]. In SemCom, this adaptability is crucial for addressing multiple heterogeneous security threats. The semantic codecs can be composed of multiple experts to resist various security vulnerabilities. A gating network is incorporated into the SemCom system to optimize the selection of these experts. This network intelligently chooses the most appropriate expert or combination of experts based on the current security landscape and the transmitter's privacy requirements [13]. This dynamic selection process ensures that the semantic encoding and defense strategies are both effective and aligned with the specific security needs of the communication, thereby improving overall trustworthiness.

By leveraging specialized experts and a strategic gating mechanism, this approach intensifies the ability of SemCom systems to safeguard sensitive information and maintain effective communication.

In response to heterogeneous security and privacy challenges, this article pioneers an MoE-based trustworthy SemCom system optimized to improve the security and reliability of 6G networks. The main contributions of this article are as follows:

- To meet the trustworthy security and privacy requirements in 6G networks, we discuss the promising applications and derived challenges of SemCom, which motivate our use of MoE to invoke appropriate defense schemes for addressing multiple heterogeneous threats simultaneously.
- The proposed MoE-based SemCom system can adaptively select the semantic codec according to user-defined security requirements using a gating network. Additionally, our designed SemCom system is extensible, allowing seamless integration of emerging security experts to achieve sustainable trustworthiness for 6G networks.
- We provide a case study of vehicular message sharing to show the practicality of our proposed scheme. It considers multiple heterogeneous wireless communication threats including source attacks, channel attacks, detection attacks, and eavesdropping attacks. Numerical results show that the proposed SemCom system can achieve high accuracy while meeting 6G security and privacy requirements in the face of multiple attacks.

Semantic Communication in 6G Network

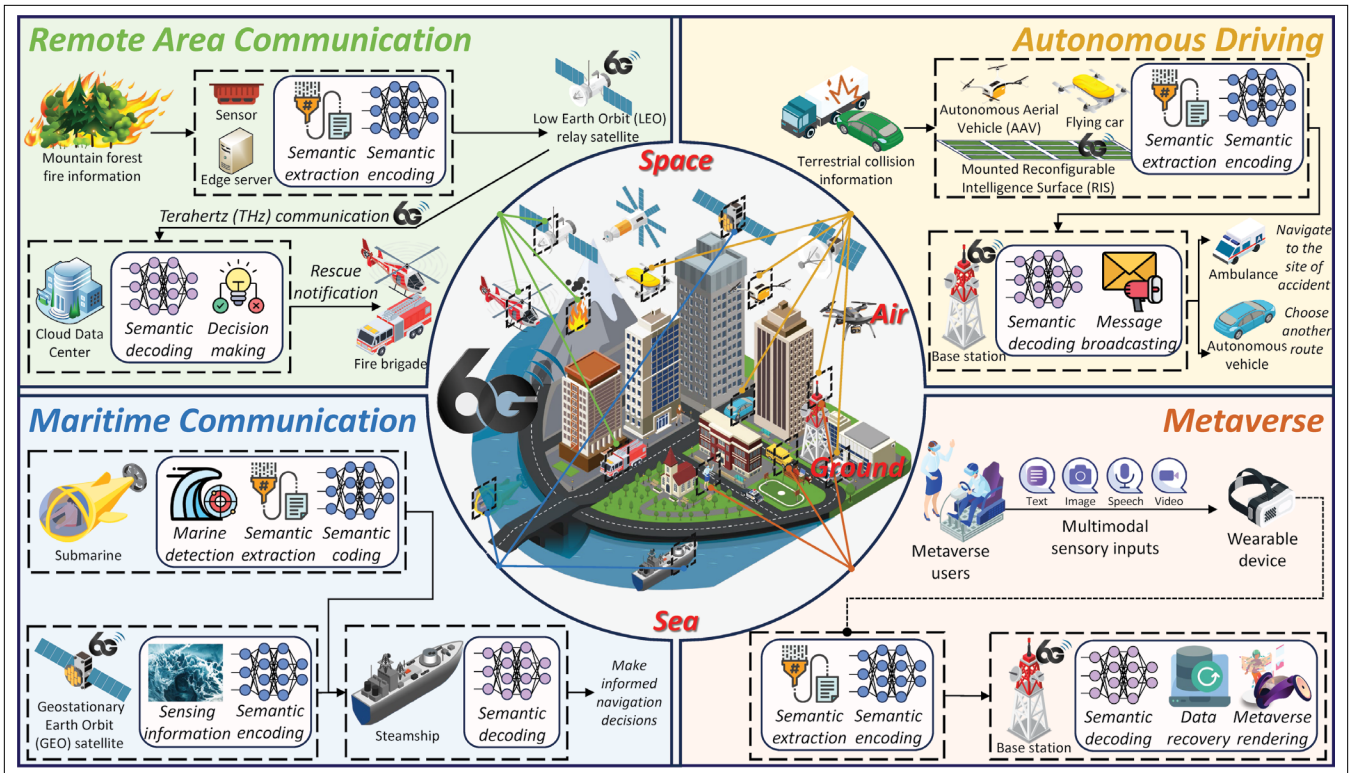
Review of Semantic Communication

SemCom concentrates on extracting the semantic contents of the source message at the transmitter and interpreting their meanings at the receiver. By exchanging semantic information at both ends, SemCom enables the reconstruction of source messages or direct execution of downstream tasks with tolerance for transmission errors [1]. In a DL-based SemCom system, a semantic encoder at the transmitter effectively encapsulates the essential meaning behind messages while a semantic decoder at the receiver reconstructs the transmitted semantic information into an understandable form. Various DL models can be employed as semantic codecs to process different types of semantic information. For example, the Transformer exhibits high proficiency in Natural Language Processing (NLP) tasks and is innovatively used in textual SemCom systems [14]. Besides, SemCom building upon other DL models such as Variational AutoEncoders (VAEs), Generative Adversarial Networks (GANs), and diffusion models can extract features from different data modalities like images and audio, thereby enhancing the efficiency of message transmission over wireless channels [1].

Semantic Communication Applications in 6G Networks

The 6G networks have remarkable features such as space-air-ground-sea seamless coverage, extensive virtualization, and ubiquitous intelligence [15]. Considering the great potential of SemCom to enable the above features, we present four potential application scenarios of SemCom in space-air-ground-sea integrated 6G networks, while presenting a comprehensive comparison between SemCom and traditional communication techniques, as depicted in Fig. 1.

- **Remote Area Communication:** SemCom can enhance disaster monitoring in remote areas to facilitate emergency response by relevant institutions. For example, if a forest fire occurs deep in the mountain with poor communication conditions, edge servers can extract the essential fire information (e.g., location) detected by the adjacently installed sensor, and then forward it to the neighboring data center after the relay of Low Earth Orbit (LEO) satellite through Terahertz (THz) communication [1]. The cloud data center can promptly decode the semantic representation to acquire the original complete information and then notify the nearest fire brigade to the scene of the fire.



Communication techniques	Principles	Advantages	Disadvantages	Application scenarios
Traditional communication	• **Raw Data Transmission:** Transmit raw bits or uncompressed information, focusing on accuracy and fidelity rather than context. • **Error Correction:** Ensures data integrity by various error detection and correction methods.	• **Mature and Widely Used**: Proven infrastructure and widely deployed technology. • **High Data Fidelity**: Preserves the full details of transmitted data.	• **High Bandwidth Demand**: Requires significant bandwidth for high-volume or real-time applications. • **Latency Issues**: Not suitable for time-sensitive tasks. • **Energy Inefficient**: Requires more energy to transmit large amounts of raw data.	***Universal applications in 4G and 5G networks*** • **High-fidelity Voice Communication:** Telephony, Voice over Internet Phone • **Accurate Data Transfer:** Hyper Text Transmission, File Transfer • **Signal Broadcasting:** TV, Radio, Satellite Transmission
Semantic Communication (SemCom)	• **Semantic Information Transmission:** Transmits only the relevant information rather than raw data. • **Contextual Encoding:** Focuses on transmitting essential data that captures the "meaning" to increase efficiency.	• **Bandwidth Efficient:** Transmits only essential semantic data, significantly reducing bandwidth usage. • **Reduced Latency and Energy:** Reduces delays and energy consumption by minimizing the amount of data transmitted.	• **Loss of Detailed Information:** Sacrifices some details when transmitting the essential content. • **Specialized Demand:** Requires new techniques and infrastructure that may not be widely available.	***Emerging applications in B5G and 6G networks*** • **Immersive and Interactive Communication:** Metaverse, Telesurgery in healthcare systems • **Ultra Low-Latency Environments:** Autonomous driving, IoT systems • **Bandwidth-Constrained Environments:** Remote area and Maritime communication

FIGURE 1. Overview of semantic communication in 6G networks.

- **Autonomous Driving:** In an intelligent transportation system, autonomous vehicles need to keep connected with other vehicles and roadside infrastructures, e.g., RoadSide Units (RSUs), to enhance contextual awareness of road conditions for safe driving [11]. SemCom fosters connected vehicles to share road information with lower latency and higher efficiency. For instance, when low-attitude vehicles, e.g. Autonomous Aerial Vehicles (AAVs), find a traffic accident such as a terrestrial collision during the flight, they can efficiently convey the coded semantic information of the accident, even in urban areas with abundant tall buildings.
- **Maritime Communication:** The harsh oceanic wireless environment challenges maritime communications due to poor channel conditions causing bit distortion [1]. SemCom mitigates this by conveying the meaning of messages rather than bit streams. For example, a submarine detecting an unknown marine creature with sonar can semantically encode and transmit the information to a targeted vessel (e.g., a steamship) via an acoustic semantic communication system. Meanwhile, a Geostationary Earth Orbit (GEO) satellite can use remote sensing technology to monitor the state of the corresponding sea area and transmit it to the steamship. By decoding both semantic sources, the steamship can comprehensively understand the maritime situation, thereby making informed navigation decisions.
- **Metaverse:** SemCom is expected to play a transformative role in the operation of metaverses [3], [6]. By capturing vital

Attack	Tampering		Eavesdropping	
Adversary	Source noise	Channel noise	Warden	Eavesdropper
Goal	1. Cause the receiver to get wrong information 2. Tamper the results of downstream tasks		Obtain the transmitter's information	Obtain the contents of source messages
Knowledge	○: Structure and weights of semantic encoder ●: Inputs and outputs of semantic encoder	○: Structure and weights of semantic codec ●: Inputs and outputs of semantic codec	–	Structure and weights of semantic decoder
Action	Add noise into the source	Add noise into the channel	Detect the occurrence of transmitting activities	Decode the semantic information
Countermeasures	Adversarial training [6]	Adversarial training [7]	Covert communication [9]	Encryption [10]

The ○ represents white box attack, while the ● means black box setting.

TABLE 1. Comparison of attacks and defenses for semantic communication in 6G networks.

semantic information for transmission, SemCom dramatically reduces communication latency, thereby guaranteeing users' immersive experiences in metaverses. For instance, wearable devices, e.g., VR glasses, can collect multimodal sensory inputs, such as text, image, speech, and video from metaverse users. Due to the limited computing resources of wearable devices, computation-intensive tasks like rendering avatars and virtual spaces should be offloaded to edge servers for further processing [11]. To this end, the raw data are transmitted to the base station via a SemCom system [6]. With the recovered user data, the base station can render a hyper-realistic metaverse and provide remarkable metaverse services back to users.

Threats

As discussed above, SemCom has numerous potential applications that can shape our future lifestyle. Therefore, it is necessary to investigate the security issues and defenses associated with SemCom. As shown in Table 1, attacks against SemCom can be divided into two categories based on the goals of attackers, namely, tampering information and eavesdropping information. Tampering can occur at the source and the channel, while eavesdropping is divided into two steps: detection of communication status and acquisition of source information. This section introduces these threat models along with existing defense schemes.

Tampering Attack From Source. *1) Threat model:* By injecting adversarial semantic noise into the source, an attacker can make the semantic decoder fail to restore the source information accurately, leading to incorrect information acquisition at the receiver and erroneous outcomes for downstream tasks. In the white-box mode, the attacker fully grasps the model parameters and weights of the semantic codec, thereby accurately adjusting the gradient information to obtain a minimum semantic noise that causes the semantic codec to produce wrong results. In the black-box mode, the attacker can only query the input and output of the codec a finite number of times to achieve the same goal without gradient information [6]. Note that a source attack does not imply that the transmitter is the attacker, as an attacker could post noisy images in the real world (e.g., post-it notes on a traffic sign) that the transmitter unknowingly captures and transmits, thus unexpectedly triggering the attack. *2) countermeasures:* The defender simulates an adversary that injects noise into the source, employing adversarial training to enhance the robustness of the semantic encoder-decoder to the source noise.

Tampering Attack From Channel. *1) Threat model:* An attacker can introduce semantic noise into the channel to achieve the same effect as tampering at the source. However, generating the semantic noise for the channel requires the attacker to possess more knowledge. In the white-box mode, the attacker needs to know the model parameters of both the semantic encoder and the semantic decoder, while even in the black-box mode, the attacker is required to be able to query the complete communication input and output. *2) countermeasures:* The defender simulates an adversary that injects noise into the channel, employing adversarial training to enhance the robustness of the semantic encoder-decoder to the channel noise.

Eavesdropping Attack by a Warden. *1) Threat model:* The goal of a warden is to detect the presence of wireless transmission activity to decipher the location of the communicator or engage in other harmful behaviors. The warden collects signals in the wireless channel at a certain frequency to detect communication activity, which requires the ability to acquire signals in the open wireless channel and analyze the traffic using algorithms. *2) countermeasures:* The primary defense scheme is to use covert communication techniques, which involve setting a friendly jammer to emit noise into the communication amplifier to mask the real communication activity. In SemCom, users can also increase the semantic compression rate to reduce the probability of communication being detected [9].

Eavesdropping Attack by an Eavesdropper. *1) Threat model:* Following the warden, an eavesdropper aims to obtain the semantic information transmitted in the channel and correctly decode it to obtain the contents of source messages. This requires the eavesdropper not

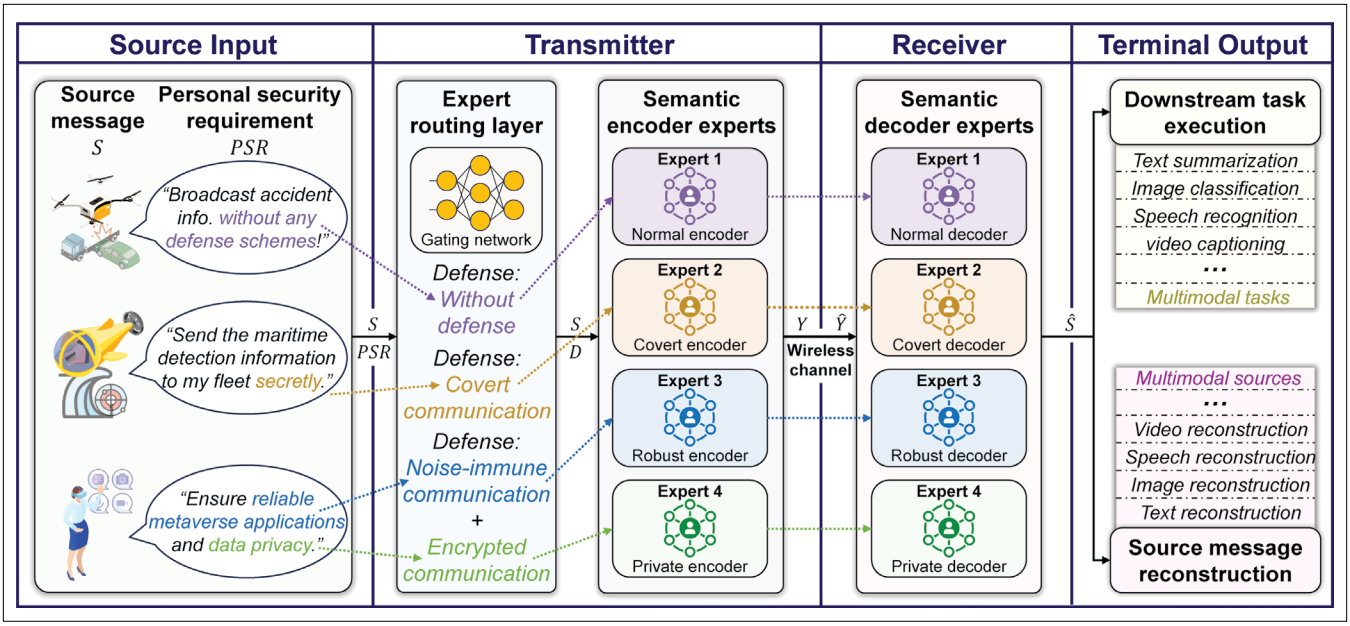


FIGURE 2. A general Mixture-of-Experts (MoE)-based Semantic Communication (SemCom) system building upon user-defined security requirements toward trustworthy 6G networks. S is the source information, PSR is personal safety requirements, D is the decision of the access control network, Y is the semantic information, $\hat{Y}$ is the semantic information received by the recipient, and $\hat{S}$ is the source information recovered by the recipient.

only to capture the signal in the wireless channel but also to possess a similar semantic decoder. Using the physical layer technologies, the eavesdropper recovers the source contents with an illegally obtained decoder. *2) countermeasures:* Encryption is an effective defense scheme to prevent the eavesdropper from correctly decoding the signal in the channel [10]. It generates a pair of keys held by the legitimate transmitter and receiver. The transmitter mixes the source information with the key and encodes it, and the receiver uses the corresponding key to decode it correctly.

Although there are corresponding defense schemes for each attack, actual communication scenarios often face multiple heterogeneous security threats simultaneously. Existing methods, including training codecs with specialized techniques, additional encryption operations, or compression operations, cannot simply be merged to address this more complex and realistic scenario. Therefore, it is necessary to explore a new approach that seamlessly integrates various defense schemes into a SemCom system to support trustworthy 6G networks.

Mixture-of-Experts (MoE)-Based Semantic Communication for Trustworthy 6G Networks

Although existing research works have each proposed a defense scheme to prevent a certain attack in SemCom systems, these approaches cannot simultaneously defend against multiple types of malicious attacks according to users' specific security requirements during communications, e.g., covertness and robustness. This limitation is particularly problematic in real-world scenarios. Therefore, to enable the practical application of SemCom, it is imperative to design an effective SemCom system to preserve telecommunication security under multiple heterogeneous threats for trustworthy 6G networks.

Although there are corresponding defense schemes for each attack, actual communication scenarios often face multiple heterogeneous security threats simultaneously.

Mixture-of-Experts (MoE) Network Model

The MoE network typically involves a gating network and several specialized trained models (i.e., experts), which can perform diverse machine learning tasks by selectively activating a subset of experts according to model inputs [12]. This simple network architecture caters to the demands for enhancing the trustworthiness of SemCom with a reasonable computational cost. In addition, the MoE can easily adapt to new datasets and AI tasks by adding or re-training experts without overhauling the entire network. With this scalable feature, an MoE-based SemCom system can retain normal operation by incorporating a newly trained expert whenever a novel security issue arises.

An MoE-Based Semantic Communication System for Trustworthy 6G Networks

As shown in Fig. 2, we propose an MoE-based SemCom system to handle users' diverse security requirements in 6G networks. The proposed SemCom system consists of a gating network, multiple semantic encoder experts for each modality at the transmitter end, and corresponding semantic decoder experts at the receiver end. The process of handling the source message and requirements in the MoE-based SemCom system is described below.

- **Step1:** The transmitter takes the source message S and the user's personal security requirements PSR as inputs to the gating network.

- **Step2:** The gating network outputs the decision D about the selection and execution order of the experts based on source message and requirements.
- **Step3:** According to the decision D, the transmitter calls the corresponding experts (semantic encoders and defense methods) to encode the source message S to obtain the semantic information Y.
- **Step4:** After channel encoding, wireless channel transmission, and channel decoding, the receiver obtains the semantic information $\hat{Y}$.
- **Step5:** By calling the corresponding decoder experts, the receiver recovers the source message $\hat{S}$.

To satisfy various user-defined security requirements, the data inputs from users of the proposed MoE-based SemCom system involve source messages and users' personal security requirements. By analyzing user objectives from the personal security requirement, the gating network can orchestrate the selection of experts to conduct security-oriented or privacy-preserving semantic encoding and decoding during SemCom. For example, users seek to immersively experience metaverse applications without privacy leakage [11]. This demands that during message transmission, the sensitive sensory data inputs (i.e., the source message S) should be anti-tampering and encrypted (i.e., the personal security requirement PSR). Based on PSR and S, the gating network makes decisions D to sequentially call the trained robust encoder and privacy encoder to extract the semantic information Y credibly [13]. The extracted semantic information is then processed as $\hat{Y}$ by the channel encoder, wireless channel, and channel decoder. Then the privacy decoder and robust encoder at the receiver process the $\hat{Y}$ to recover $\hat{S}$ for the output. Since both robust codecs and privacy codecs are invoked, the communication process avoids message tampering and eavesdropping. The output $\hat{S}$ can be applied in various tasks, such as source message reconstruction and downstream task execution [4].

Gating Network and Experts Architectures and Training Method

1) Gating Network: The key component of the MoE network is the gating network, which routes the source message to the proper experts. Here, we deploy a MultiLayer Perceptron (MLP) at the transmitter as the gating network to select one or more semantic codec experts based on users' security requirements [12]. The user's security and privacy requirements are embedded as a vector PSR as part of the input to the gating network, with the source message S to be transferred constituting the other part. The output of the gating network is a vector representing the decision D. During training, we randomly generate a dataset of uniformly distributed user requirements, and each requirement corresponds to a correct expert selection as the label. A user may have multiple requirements at the same time, corresponding to multiple labels. The loss function of the gating network comprises the selection loss and selection penalty of semantic codec experts, guiding the parameter updates of the MLP. The selection loss ensures that the gating network correctly routes the input to specialized experts to enhance task accuracy and meet user-defined security requirements. The selection penalty ensures that the gating network does not select unnecessary experts, reducing computing costs. Note that the gating network is trained after the expert's training, and the experts' model parameters are frozen during the gating network training. Therefore, when new experts are added, only the gating network needs to be fine-tuned, eliminating the need to retrain the entire system [12]. This architecture provides greater scalability for the SemCom system.

2) Normal Codecs: The architecture of a normal codec consists of a semantic encoder, a semantic decoder, a channel encoder, and a channel decoder, as this encoder-decoder architecture can extract critical semantic features to improve communication efficiency [2], [14]. The semantic and channel encoder is deployed at the transmitter end, while the semantic and channel decoder is deployed at the receiver end. The structure of normal codecs is different for different communication tasks. For example, for text transfer tasks, the Transformer model is used as a semantic codec [14]. For image transmission and classification tasks, Resnet is used as a semantic codec [6]. Regardless of the specific structure of the codec, the transmitter first feeds the S into the semantic encoder to extract the semantic information Y. Y is then sent to the receiver after passing through the channel encoder. Upon receiving the information, the receiver performs channel decoding to recover the semantic information $\hat{Y}$ and semantic decoding to recover the source message $\hat{S}$. By iteratively updating the model parameters of the semantic codec and channel codec based on the communication loss, we can train a normal codec expert.

3) Covert Codec: In traditional communications, a friendly jammer is usually set up to send a jamming signal to the transmitter to prevent the warden from detecting that communication is taking place. In SemCom, the semantic compression rate can be increased to improve the covert rate [9]. The warden detects communication activities in the channel at a certain frequency. If a communication endures for an extended duration, each detection it encounters increases the likelihood of successful detection by the warden. Increasing the compression rate can reduce the number of detection times that the communication undergoes, thereby reducing the warden's detection success rate and achieving covert communication. Therefore, after obtaining the semantic information extracted by the semantic encoder, the covert codec further compresses it using a compressor. We utilize an autoencoder with an adjustable compression ratio as the compressor, where the encoder is deployed after the semantic encoder and the decoder is deployed before the semantic decoder. In the training phase, the autoencoder updates its parameters using the Mean Square Error (MSE) between the input and output as the loss.

4) Robust Codec: The Semantic Distance Minimization (SDM)-based training method is often used to combat tampering attacks from source

[6]. SDM uses adversarial training to reduce the distance between the semantic information of identical categories of images as much as possible. During training, SDM generates an online adversarial example for each benign example by adding semantic noise while updating the model parameters to make their semantic distance closer. Therefore, the semantic codec is largely immune to semantic noise. Although attacks from the source and channel are launched from different locations and in different ways, they both achieve the purpose of tampering with communication content by adding semantic noise. Therefore, the SDM is effective against tampering attacks from both the source and the channel.

5) Private Codec: The effective way to prevent information from being intercepted by eavesdroppers is to encrypt it [10]. The transmitter and receiver each hold a secret key, with the transmitter concatenating the information to be transmitted with the key before semantic encoding. The legitimate receiver uses the key and semantic decoder to correctly recover the information, while malicious eavesdroppers, lacking the corresponding key, cannot recover the information even if they possess the same semantic decoder as the receiver. The privacy codec adopts adversarial training, simulating an eavesdropper holding the same semantic decoder, a pair of legitimate transmitter and receiver holding a key pair. The legitimate transmitter-receiver updates the model parameters to minimize the legitimate receiver's loss, while the simulated eavesdropper also attempts to minimize its loss. However, due to the advantage of the legitimate receiver holding the key, its decoder model can gradually converge, while that of the simulated eavesdropper cannot.

In this section, we detail the architecture, training methods, and advantages of the proposed MoE-based SemCom system. However, calling multiple defense schemes simultaneously may result in a loss of communication accuracy. To explore its effects in real-world scenarios, we present a case study in the next section to evaluate its performance in a vehicular SemCom system.

Case Study

To evaluate the performance of the proposed MoE-based SemCom system in real-world scenarios, we conducted a case study on vehicle communications, in which the detailed system model, experimental setup and simulation results are described below.

System Model

In the case study, moving vehicles utilize semantic communication to exchange images captured by vehicular sensors, supporting applications such as autonomous driving and vehicular metaverses [6]. The environment includes different types of attackers attempting to tamper with information or eavesdrop on user privacy. The proposed MoE-based SemCom system, comprising a gating network, semantic codecs, and defense mechanisms, is deployed in the vehicles to counter these heterogeneous attacks. In the system, vehicular users customize their security and privacy requirements based on the contents being transmitted for the gating network to make decisions and engage appropriate security experts. Specifically, We consider that vehicles use SemCom to transmit images and classify them to guide driving decisions. The transmitter uses the SemCom system to encode and send the acquired image, while the receiver decodes it and completes classification tasks [6]. During the communication process, we implement the four attacks mentioned above. The proposed SemCom system includes corresponding experts to resist the attacks and a gating network to precisely call these experts. Since the communication task is image classification, a Resnet20 network is used as the normal codec [6]. The loss function for training normal codecs uses cross entropy as a classification loss.

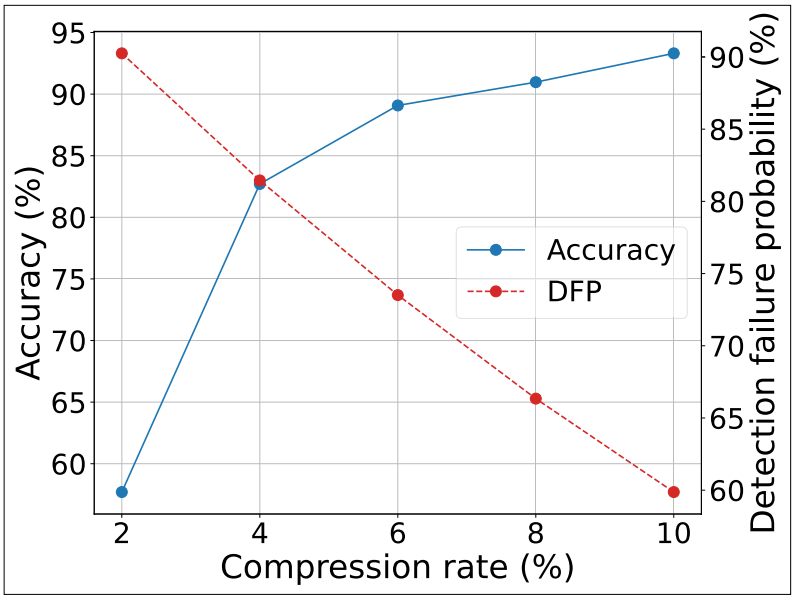


FIGURE 3. The accuracy performance and the DFP when adding covert codec experts with different compression rates to the MoE-based system.

Experimental Setup

We consider a traffic sign recognition task to evaluate the performance of the proposed MoE-based SemCom system. The dataset used in the simulation experiments is the GTSRB[1], which consists of over 50,000 images with resolutions ranging from 15 × 15 to 250 × 250 pixels, divided into 43 classes. During the performance evaluation of the MoE-based SemCom system, Additive White Gaussian Noise (AWGN) is added to the channel with a Signal Noise Ratio (SNR) between 0 and 12 [14]. For the covert codec, the detection error probability, covert rate, and bandwidth are set to 95%, 100 bit/s/Hz, and 5 MHz, respectively. Besides, the warden monitoring the open wireless channel is assumed to detect communication activity at a frequency of twice per second [9].

The primary performance metric for a task-oriented SemCom system is the classification task accuracy. Since the goal of tampering attacks is to tamper with the results of classification tasks, classification accuracy is also used to measure the effect of tampering attacks. In particular, when the system suffers from an eavesdropping attack by a warden, the Detection Failure Probability (DFP) [9] metric is utilized to assess the effectiveness of covert communication in our

[1] https://benchmark.ini.rub.de/

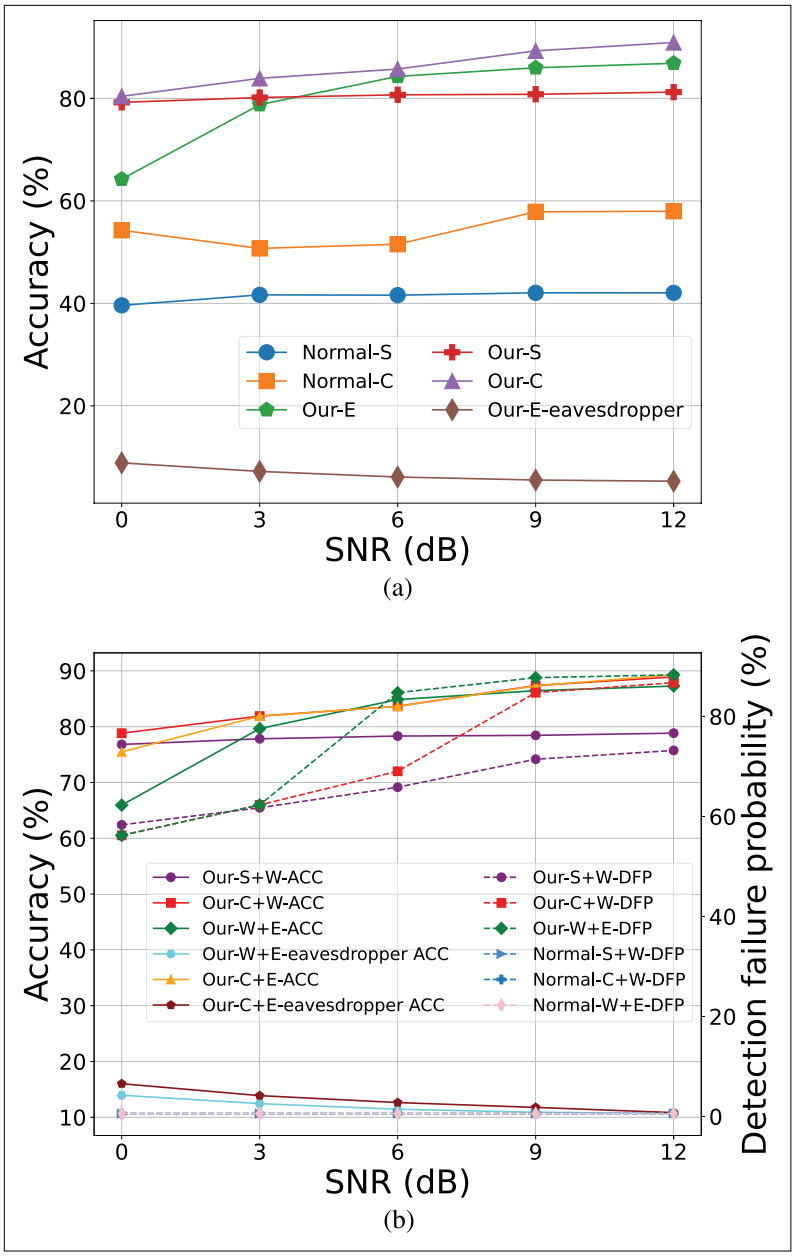


FIGURE 4. The performance of proposed MoE-based SemCom system facing single and multiple heterogeneous attacks. (S: Source noise; C: Channel noise; E: Eavesdropper; W: Warden; ACC: Accuracy; DFP: Detection failure probability). a) Single attacks. b) Multiple heterogeneous attacks.

proposed method. Meanwhile, the accuracy of an eavesdropper is also considered to evaluate the efficacy of semantic information encryption in the proposed system when facing an eavesdropping attack by an eavesdropper.

Simulation Results

Fig. 3 explores the accuracy performance and the DFP when adding covert codec experts with different compression rates to the MoE-based system. The numerical results show that the increase of compression ratio can significantly increase DFP, but the cost is that the communication accuracy will be weakened. Fig. 4 shows the performance of the proposed MoE-based SemCom system when the system suffers attacks. Fig. 4(a) shows the situation when it suffers a single attack, while Fig. 4(b) shows the cases of multiple heterogeneous attacks. Various performance indicators prove that the proposed MoE-based SemCom system can adaptively select appropriate experts to complete communication tasks based on the user's security and privacy requirements. Numerical results show that the MoE-based SemCom system can effectively resist single attacks with minimal accuracy loss.

As shown in Fig. 4(a), the MoE-based SemCom maintains a communication accuracy of about 80% under various SNR ratio conditions, while the accuracy of the normal method drops to about 40% under source noise. In addition, even in the absence of attack interference, the MOE-based approach still has higher accuracy compared to the normal method. Similarly, in the face of channel noise, the MoE-based method can maintain an accuracy of more than 80% and increase to 90% with the improvement of channel quality. The accuracy of the general method falls below 60% in the face of the channel noise. In the face of an eavesdropper, the legal receiver can obtain high communication accuracy, especially at a high SNR, while the eavesdropper can hardly decode the message content.

Fig. 4(b) illustrates the performance of the proposed approach against both source noise and a warden, showing that the MoE-based method achieves both high communication accuracy and DFP. In the face of channel noise and wardens, the proposed approach is equally advantageous. Moreover, under high SNR conditions, the MoE-based approach can achieve similar performance when facing multiple attacks as when facing a single type of attack. Under the simultaneous attack of the warden and the eavesdropper, the MoE-based approach makes the eavesdropper achieve a very low precision parallel while maintaining the accuracy of the legal receiver and the DFP compared to the normal approach. The proposed method can maintain stable communication accuracy and confidentiality in the face of channel noise and eavesdroppers. In conclusion, the MoE-based SemCom solves multiple heterogeneous attack scenarios that have never been considered by existing methods.

Future Directions

Input Simplification for MoE-Based SemCom

The proposed MoE-based SemCom requires users to input their security and privacy requirements to tailor decision-making processes accordingly. Some users anticipate that these requirements will be addressed unconsciously, requiring minimal explicit guidance. For instance, if source messages include location details, the gating network autonomously selects the privacy codec without explicit user directives. By enabling automated decision-making based on user inputs and content characteristics, such systems can adeptly address security requirements and other requirements, such as multimodal SemCom [4].

Channel Estimation for MoE-Based SemCom

Accurate channel estimation is crucial for maintaining the integrity of a SemCom system, particularly as error rates of core semantic meanings have a greater impact. As SemCom

evolves from traditional bit-based communication, a prolonged phase of hybrid bit-semantic communication is foreseen. The amalgamation of various semantic codec specialists and the hybrid nature of bit-semantic exchanges present hurdles in accurately estimating Channel State Information (CSI). Consequently, the pursuit of an effective and reliable channel estimation technique tailored for MoE-based SemCom emerges as a crucial avenue for future exploration in trustworthy 6G networks.

Resource Allocation for MoE-Based SemCom

The utilization of SemCom in resource-limited environments like the Internet of Things (IoT) is prevalent due to its capacity to notably diminish communication overhead. In MoE-based SemCom, each additional expert amplifies computational demands, augments storage requirements, and extends communication latency. In certain scenarios, prioritizing reduced energy consumption may warrant relinquishing unnecessary security measures. Moreover, optimizing storage resources by relocating seldom-used experts to the network edge rather than locally presents an avenue for efficiency. Consequently, the investigation into resource allocation assumes significant importance. Exploring the optimal approach for balancing constrained resources and security prerequisites for individual devices emerges as a pertinent future direction.

Conclusion

In response to the multi-source heterogeneous attacks existing in wireless communication environments, this article advocates integrating Mixture-of-Experts (MoE) into Semantic Communication (SemCom) to enhance the trustworthiness of 6G networks from the aspects of security and privacy.

By incorporating MoE into SemCom, the gating network can dynamically cater to users' varying security and privacy requirements of users by allocating suitable semantic codecs and security measures for each communication instance. To illustrate the efficacy of this integrated system, a case study focusing on vehicular semantic communications with four specific security threats is conducted. The results highlight the effectiveness of our approach in bolstering security and privacy in 6G networks, thereby laying a strong foundation for trustworthy network infrastructures. Finally, the potential future contributions of MoE for SemCom in advancing 6G networks are discussed, providing insights into the continued impact of MoE in promoting future communication paradigms.

The results highlight the effectiveness of our approach in bolstering security and privacy in 6G networks, thereby laying a strong foundation for trustworthy network infrastructures.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62102099 and Grant U22A2054; in part by the Guangzhou Basic Research Program under Grant 2023A04J1699; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2023A151514 0137; in part by the National Research Foundation, Singapore, and Infocomm Media Development Authority under its Future Communications Research and Development Programme, Defence Science Organisation (DSO) National Laboratories through the AI Singapore Programme under Grant FCP-NTU-RG-2022-010 and Grant FCP-ASTAR-TG-2022-003; in part by the Singapore Ministry of Education (MOE) Tier 1 under Grant RG87/22; in part by the NTU Centre for Computational Technologies in Finance (NTU-CCTF); and in part by Seitee Pte Ltd.

References

[1] H. Peng et al., "Semantic communication in non-terrestrial networks: A future-ready paradigm," *IEEE Netw.*, vol. 38, no. 4, pp. 119–127, Jul. 2024.
[2] Y. Tan, J. Liu, and J. Wang, "5G end-to-end slice embedding based on heterogeneous graph neural network and reinforcement learning," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 3, pp. 1119–1131, Jun. 2024.
[3] J. Wang et al., "Semantic-aware sensing information transmission for metaverse: A contest theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 22, no. 8, pp. 5214–5228, Aug. 2023.
[4] A. Li et al., "Cross-modal semantic communications," *IEEE Wireless Commun.*, vol. 29, no. 6, pp. 144–151, Dec. 2022.
[5] H. Xie et al., "Task-oriented multi-user semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 9, pp. 2584–2597, Sep. 2022.
[6] J. Kang et al., "Adversarial attacks and defenses for semantic communication in vehicular metaverses," *IEEE Wireless Commun.*, vol. 30, no. 4, pp. 48–55, Aug. 2023.
[7] G. Nan et al., "Physical-layer adversarial robustness for deep learning-based semantic communications," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, pp. 2592–2608, Jun. 2023.
[8] Q. Hu et al., "Robust semantic communications with masked VQ-VAE enabled codebook," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 8707–8722, Dec. 2023.
[9] H. Du et al., "Rethinking wireless communication security in semantic Internet of Things," *IEEE Wireless Commun.*, vol. 30, no. 3, pp. 36–43, Jun. 2023.
[10] X. Luo et al., "Encrypted semantic communication using adversarial training for privacy preserving," *IEEE Commun. Lett.*, vol. 27, no. 6, pp. 1486–1490, Jun. 2023.
[11] X. Luo et al., "Privacy attacks and defenses for digital twin migrations in vehicular metaverses," *IEEE Netw.*, vol. 37, no. 6, pp. 82–91, Nov. 2023.
[12] N. Shazeer et al., "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," 2017, *arXiv:1701.06538*.
[13] H. Du et al., "Mixture of experts for intelligent networks: A large language model-enabled approach," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, May 2024, pp. 531–536.
[14] H. Xie et al., "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, pp. 2663–2675, 2021.
[15] J. Wang et al., "Intelligent network slicing for B5G and 6G: Resource allocation, service provisioning, and security," *IEEE Wireless Commun.*, vol. 31, no. 3, pp. 271–277, Jun. 2024.

Biographies

Jiayi He (jiayihe@ieee.org) received the B.Eng. degree from Shandong University, Jinan, China, in 2021. He is currently pursuing the M.S. degree with the School of Automation, Guangdong University of Technology, China. His research interests include semantic communications, security and privacy, and generative AI.

Xiaofeng Luo (gdutxiaofengluo@163.Com) received the B.Eng. degree from the Guangdong University of Technology, China, in 2023. He is currently pursuing the M.S. degree with the School of Automation, Guangdong University of Technology. His research interests include privacy protection, edge intelligence, and metaverse.

Jiawen Kang (kavinkang@gdut.edu.cn) received the Ph.D. degree from the Guangdong University of Technology, China, in 2018. He was a Post-Doctoral Researcher at Nanyang Technological University, Singapore, from 2018 to 2021. He is currently a Professor with the Guangdong University of Technology.

His research interests mainly focus on blockchain, security, and privacy protection in wireless communications and networking.

Hongyang Du (duhy@hku.hk) received the B.Eng. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, in 2021, and the Ph.D. degree from the Interdisciplinary Graduate Program, College of Computing and Data Science, Energy Research Institute, Nanyang Technological University (NTU), Singapore, in 2024. He is currently an Assistant Professor with The University of Hong Kong (HKU), Hong Kong. His research interests include edge intelligence, generative AI, semantic communications, and network management.

Zehui Xiong (zehui_xiong@sutd.edu.sg) received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He was a Visiting Scholar at Princeton University and the University of Waterloo. He is currently an Assistant Professor with the Pillar of Information Systems Technology and Design, Singapore University of Technology Design, Singapore. His research interests include wireless communications, network games and economics, blockchain, and edge intelligence.

Ci Chen (ci.chen@gdut.edu.cn) received the B.E. and Ph.D. degrees from the School of Automation, Guangdong University of Technology, China, in 2011 and 2016, respectively. He was a Wallenberg-NTU Presidential Post-Doctoral Fellow at Nanyang Technological University, Singapore, and Lund University, Sweden. He is currently a Professor and the Department Chair of automatic control with the Guangdong University of Technology. His research interests include reinforcement learning for feedback control, cooperative vehicles, resilient designs, and computational intelligence.

Dusit Niyato (Fellow, IEEE) (dniyato@ntu.edu.sg) received the B.Eng. degree from the King Mongkut's Institute of Technology Ladkrabang (KMITL), Thailand, in 1999, and the Ph.D. degree in electrical and computer engineering from the University of Manitoba, Canada, in 2008. He is currently a Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. His research interests include the Internet of Things (IoT), machine learning, and incentive mechanism design.

Xuemin (Sherman) Shen (Fellow, IEEE) (sshen@uwaterloo.ca) received the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 1990. He is currently a University Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include network resource management, wireless network security, the Internet of Things, AI for networks, and vehicular networks. He is a Registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, a Chinese Academy of Engineering Foreign Member, an International Fellow of the Engineering Academy of Japan, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and Communications Society.