SCIENCE CHINA
Information Sciences

July 2022, Vol. 65 170301:1–170301:11
https://doi.org/10.1007/s11432-021-3463-9

• RESEARCH PAPER •
Special Focus on Cyber Security in the Era of Artificial Intelligence

Intelligent networking in adversarial environment: challenges and opportunities

Yi ZHAO[1], Ke XU[1,3*], Qi LI[2,3], Haiyang WANG[4], Dan WANG[5] & Min ZHU[1]

[1]*Department of Computer Science and Technology, Tsinghua University, Beijing* 100084*, China;*
[2]*Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing* 100084*, China;*
[3]*Beijing National Research Center for Information Science and Technology* (*BNRist*)*, Beijing* 100084*, China;*
[4]*Department of Computer Science, University of Minnesota Duluth, Duluth* MN 55812*, USA;*
[5]*Department of Computing, The Hong Kong Polytechnic University, Hong Kong* 999077*, China*

Received 25 August 2021/Revised 7 January 2022/Accepted 25 March 2022/Published online 21 June 2022

Abstract Although deep learning technologies have been widely exploited in many fields, they are vulnerable to adversarial attacks by adding small perturbations to legitimate inputs to fool targeted models. However, few studies have focused on intelligent networking in such an adversarial environment, which can pose serious security threats. In fact, while challenging intelligent networking, adversarial environments also bring about opportunities. In this paper, we, for the first time, simultaneously analyze the challenges and opportunities that the adversarial environment brings to intelligent networking. Specifically, we focus on challenges that the adversarial environment will pose on the existing intelligent networking. Furthermore, we investigate frameworks and approaches that combine adversarial machine learning with intelligent networking to solve the existing deficiencies of intelligent networking. Finally, we summarize the issues, including opportunities and challenges, which can allow researchers to focus on intelligent networking in adversarial environments.

Keywords intelligent networking, adversarial, attacks, defense, security

Citation Zhao Y, Xu K, Li Q, et al. Intelligent networking in adversarial environment: challenges and opportunities. Sci China Inf Sci, 2022, 65(7): 170301, https://doi.org/10.1007/s11432-021-3463-9

1 Introduction

Owing to the continuous accumulation of large-scale data and rapid development of computing resource capabilities, such as graphics processing units and field-programmable gate array, deep learning (DL)-based artificial intelligence (AI) technologies have been successful in providing human-level capabilities for a variety of tasks. For example, in computer vision, computers have been able to automatically predict abnormal behaviors that may occur in videos for early warning and avoidance of dangerous events. In fact, the large-scale accumulation of data is largely due to the development of the Internet. The widespread deployment and application of the Internet, especially the mobile Internet, can facilitate the continuous and rapid accumulation of data from various industries, enabling anyone to access a huge amount of data in arbitrary places.

With the rapid development of intelligent technologies, the Internet has provided large-scale data for intelligent technologies, and intelligent technologies have also been widely utilized in various fields of the Internet, such as data mining-based approaches for threat detection [1,2], network resource management based on deep reinforcement learning (DRL) in datacenters [3], zero-touch network slicing [4], and the comprehensive utilization of location information in mobile Internet scenarios based on DL [5]. Particularly, in today's 5G networks, more scenarios of human daily life require the support of various network technologies, e.g., Internet, cloud computing, edge computing, and Internet of Things (IoT). Emerging next-generation network technologies, such as autonomous 5G networks and zero-touch networks [6], can

* Corresponding author (email: xuke@tsinghua.edu.cn)

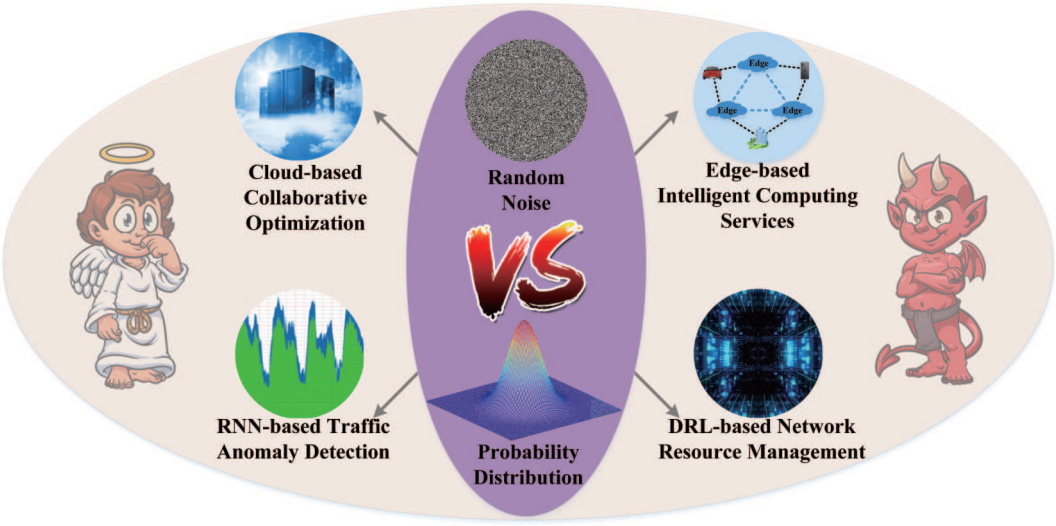

Figure 1 (Color online) Examples of intelligent networking in the adversarial environment. Is it an angel? Still the devil!

better meet the different demands under different scenarios, which undoubtedly require the support of various intelligent networking technologies.

In terms of intelligent networking technologies, existing methods with pleasant performance mainly focus on deep neural networks (DNNs) [7, 8]. However, numerous studies in areas such as computer vision and speech recognition have illustrated that DNN-based models are vulnerable to attacks with the addition of small perturbations to legitimate inputs to fool targeted models. In other words, these legitimate inputs with small perturbations can be considered one type of adversarial example, which forces existing DNN-based models to lose their original performance (e.g., reliability and accuracy) in adversarial environments, leading to serious adverse consequences. Some studies [9, 10] have found that DNN-based models face similar problems in intelligent networking fields, such as intelligent modulation recognition and intelligent spectrum sensing in wireless communication. For example, a transmitter equipped with a DNN can predict idle time slots for data transmission through its spectrum sensing results. However, Sagduyu et al. [10] proposed a novel attack method for this scenario. By injecting adversarial perturbation over the air, the attacker can fool the transmitter into making the wrong decision or failing the subsequent training. For intelligent networking, performance indicators, such as robustness and reliability, are the necessary foundations to support the widespread deployment of various intelligent networking technologies and directly affect all aspects of peoples' lives. However, few studies have focused on existing intelligent networking technologies in such an adversarial environment.

If today's AI-driven networks (e.g., autonomous 5G networks and zero-touch networks) are required to become more widely available, then the challenges of intelligent networking technologies in adversarial environments have to be fully investigated, and effective solutions should be proposed and implemented as soon as possible. In fact, while challenging intelligent networking, the adversarial environment also brings about opportunities. As illustrated in Figure 1, some intelligent networking-related technologies or applications are illustrated, including cloud-based collaborative optimization, edge-based intelligent computing services, recurrent neural network (RNN)-based traffic anomaly detection, and DRL-based network resource management.

Taking DRL-based network resource management as an example, through continuous interactions between the system and environment, the network resource management system can conduct online learning on how to perform flow scheduling and load balancing [3]. This method can reduce the cost of human intervention under the premise of ensuring service performance. In the learning process, the monitoring system has to accurately collect environmental states (i.e., traffic states); otherwise, it will cause serious accidents. However, these states are easily evolved into adversarial examples due to some random noises or probability distributions. Particularly in real-world systems, random noise is inherently inevitable. For example, due to the large traffic scale and restricted costs, network monitoring systems choose to randomly sample from the full traffic. The active discarding here can be regarded as a kind of noise. Moreover, severe communication conditions and unreliable network transmission protocols may inevitably cause loss of valuable data, which is also a kind of noise. In addition, some crafted adversarial

samples [10–12] are undoubtedly noise, and their harm is very prominent. If such issues cannot be solved, then the system can easily lose its reliability and cannot be widely deployed in the real world. The adversarial environment at this time is a devil for intelligent networking technologies. From another perspective, we can actively interfere with the system to discover potential problems. Based on the idea of game theory, a resilient autonomous network control system can be implemented via adversarial machine learning. At this time, the adversarial environment can be regarded as an angel. Because the related research is still insufficient, the pros and cons of the adversarial environment are not very clear. In other words, the adversarial environment may be not only a devil but also a valuable angel. This case is also the reason for the existence of a layer of gray shadow in Figure 1.

Overall, the adversarial environment not only poses challenges to existing intelligent networking technologies based on DNN models. It also allows existing intelligent networking technologies to obtain new opportunities to achieve even greater improvements. Investigating the challenges and opportunities of intelligent networking technologies in adversarial environments is conducive to the further development and wider popularization of AI-driven networks.

To facilitate a comprehensive understanding of existing intelligent networking technologies in adversarial environments, in this paper, we pose issues and challenges to advance the area of intelligent networking in adversarial environments. Because adversarial environments also provide opportunities for the development of intelligent networking, we present some insights for further studies. Specifically, we propose some frameworks and approaches to facilitate the integration of adversarial machine learning with existing intelligent networking technologies. Finally, we provide a comprehensive summary of the challenges and opportunities that adversarial environments bring to intelligent networking. To the best of our knowledge, this is the first paper that provides an extensive analysis of existing intelligent networking technologies in adversarial environments and proposes some valuable frameworks to pave the way for future advancements.

2 Related work

In this section, we separately review the development of related fields from the representative literature of intelligent networking, adversarial environment (e.g., adversarial attack and adversarial machine learning), and intelligent networking in adversarial environments.

Intelligent networking. Various DNN-based technologies have promoted the progress of traditional networks to intelligent networks. Moreover, intelligent networking technologies [3, 8, 13] have attracted widespread attention from academia and the industry, and have been developing rapidly. As early as the 1990s, Boyan et al. [14] proposed the utilization of reinforcement learning (RL) for flow scheduling and load balancing. However, it is difficult to implement at line rate in modern datacenters. Inspired by the great success of DRL in the field of robot control, Chen et al. [3] proposed a DRL-based automatic traffic optimization method, achieving significantly better performance than conventional RL-based methods. Similarly, through leveraging emerging DL methods, Xu et al. [13] implemented a DRL-based congestion control framework, which can enable the network to control itself according to interactions with the environment. And it has been demonstrated that the DRL-based framework is flexible and robust to highly-dynamic network environments.

In this paper, we only take the DRL-based method as an example to clarify the role of DNN-based method in promoting intelligent networking. In fact, intelligent networking based on DNN or DNN variants (e.g., DRL, RNN, and deep collaborative learning) involves a very wide range of fields, such as network intrusion detection [2] (a.k.a., malicious traffic detection [15]), online routing [16], and congestion control [17].

Adversarial attacks and adversarial machine learning. Although DL has been found to be remarkable in many scenarios where conventional shallow machine learning methods are difficult to achieve satisfactory performance, recent studies [18–20] have demonstrated that DNN-based models are vulnerable to crafted adversarial examples. For example, the crafted adversarial examples may be imperceptible to human eye, but they are possible to make conventional high-performance classification models based on DL produce incorrect classification results. Similar issues have been found in many application scenarios (e.g., network intrusion detection [21], computer vision [22], and speech recognition [23]) and intelligent models (e.g., graph neural network [24], federated learning [25], and meta learning [26]). The adversarial environment brings issues and challenges to DL, but it also paves a new path to achieve more resilient

and reliable DNN-based models [27, 28]. For example, based on game theory, Chivukula et al. [27] proposed a secure convolutional neural network (CNN), which can defend against crafted adversarial example attacks. That is, the CNN trained on adversarial examples is more resilient and reliable.

Overall, the adversarial environment (e.g., adversarial attack and adversarial machine learning) puts the existing DNN-based technologies in the scenario where opportunities and challenges coexist.

Intelligent networking in adversarial environments. Similar to other fields, DNN-based models in the field of intelligent networking [9, 10] also expose vulnerabilities and require enhanced robustness. For example, Lin et al. [9] utilized multiple gradient-based methods to generate perturbations, and added them to the original input signal to create adversarial samples, which can significantly reduce the accuracy of DNN-based modulation recognition in the field of wireless communication. As an important part of the intelligent network, the vulnerability of intelligent network intrusion detection system (NIDS) in adversarial attacks has been demonstrated via some literatures [11, 12, 21]. Through the cooperation between model extraction and saliency map, Qiu et al. [11] only modified less than 0.005% of the bytes in malicious packets to disguise legitimate packets in the IoT environment, which significantly reduces the ability of NIDS (i.e., KitSune [2]) to distinguish between legitimate and malicious activities.

In addition, the adversarial environment also brings new opportunities to the intelligent networking [12, 21, 29]. Inspired by the ability of adversarial attacks to fool DNN-based models, Hameed et al. [29] actively perturbed the channel input symbols to prevent DNN-based intruders, thereby protecting the wireless communication link. Via exploiting the inconsistency of crafted adversarial examples between NIDS inference and manifold evaluation, Wang et al. [21] established an auxiliary adversarial example detection system to weaken the negative impact of suspicious inputs on the robustness of intelligent NIDS. Han et al. [12] established an effective adversarial attack approach with the support of generative adversarial network (GAN). While fully analyzing the vulnerability of intelligent NIDS, Han et al. [12] also proposed a method of partial feature elimination to proactively enhance the robustness of NIDS.

Since the Internet has already penetrated into all aspects of human life, extensive DNN-based intelligent networking technologies are deployed in the physical world, including many safety-critical or robustness-critical environments. Although adversarial attacks have received a lot of attention in fields such as computer vision, research on the interaction between adversarial environment and intelligent networking is still in its infancy. In other words, we have seen only the tip of the iceberg with regard to the intelligent networking in adversarial environments. In our work, we investigate and summarize the challenges and opportunities of intelligent networking technologies in adversarial environments, paving the way for subsequent improvements in this area.

3 Adversarial attacks and adversarial machine learning

In this section, we summarize the essences (i.e., the goals and the approaches) of the adversarial environment (i.e., adversarial attacks and adversarial machine learning) from a global perspective, paving the way for the analysis of specific scenarios in Section 4.

3.1 Adversarial attacks

The goals. Through some specific algorithms or elaborate mechanisms to construct adversarial examples, the adversary forces the well-trained machine learning models, especially DNN-based models, to lose robustness and reliability, or trigger the leakage of privacy data, resulting in relevant models not being able to provide regular services.

The approaches. Deep learning is mainly composed of a large number of neurons, which belongs to biologically-inspired approach. However, the current understanding of neurons in the human brain is only the tip of the iceberg. Although DL has achieved good results, therefore, it is difficult to quickly and accurately handle various differential problems like the human brain. In terms of adversarial attacks, adversarial examples may refer to fake data instances that are carefully crafted by adversaries. For example, Usama et al. [30] utilized crafted noise to fool the modulation classifier. In addition to some intuitive changes, DNN can also be used to generate fake data for attack. Sagduyu et al. [10] applied DNN to enable the adversary to learn behaviors of the transmitter, thereby launching various poisoning attacks. In terms of attack categories, adversarial attacks can be divided into multiple categories due to differences in attack methods or prerequisites. As illustrated in Figure 2 [12, 28, 31–35], we provide some representative examples of adversarial attack classification.

Figure 2 (Color online) Examples of adversarial attack classification based on different indicators. Specifically, the adversarial attacks involved in this figure are associated with many real-world scenarios. This paper provides some specific cases, including black-box attack [12], grey-box attack [31], and white-box attack [32]; iterative attack [33] and one-shot attack [32]; targeted attack [28] and non-targeted attack [34]; gradient-based attack [32], GAN-based attack [12], and optimization-based attack [35].

In fact, even if no adversary actively designs adversarial examples, adversarial attacks are ubiquitous. The main reason is that the current DNN-based model is based on specific or partial data training. It is difficult to ensure that the dataset can cover all the samples in the real world. Coupled with the errors or differences in the various data acquisition tools, models that perform well during the training phase may still expose various problems similar to those that are attacked.

In addition to various attacks that invalidate the DNN-based models, there are also application-level attacks. These attacks throw a large number of fake samples to the DL model, and then, they can obtain the privacy information of the real data used by the training model. Thus, the issue of privacy breaches can also be regarded as an attack. A similar situation limits the source of the training data. However, data is one of the most important factors for DL (the other factor is computing power), which is undoubtedly very serious.

3.2 Adversarial machine learning

The goals. Based on game theory or mechanism design, and analysis for specific scenarios, adversarial machine learning focuses on promoting machine learning models to defend against attacks from adversarial examples, improving the robustness and reliability, as well as protecting privacy data.

The approaches. In terms of DL, the mechanism inside the neurons is not completely known. Therefore, DL models have many potential vulnerabilities, which are exploited by adversarial attacks. One of the most extensive adversarial machine learning approaches is achieved by an auxiliary neural network that confronts the major neural network. Based on the idea of game theory, both parties continuously improve their performance without knowing the internal principle of their own vulnerabilities. Once the game equilibrium is reached, some potential vulnerabilities in the neural network are eliminated, making DL more resilient and reliable.

In addition to confrontation based on the idea of game theory, it is also possible to characterize adversarial examples based on a specific scene. In terms of characterizing adversarial examples, the temporal dependency-based method has been proposed for audio adversarial examples [36], which can effectively improve the ability of the discriminator against adversarial examples. In fact, intelligent networking can also utilize similar unique properties of data to improve robustness. For example, there is the temporal dependency between traffic data in the field of intrusion detection. In the case of adversarial attacks, the victim should exploit the unique properties when designing the corresponding DNN-based models, and then identify adversarial samples.

In summary, whether based on the relevant theories of game theory or not, adversarial attack and adversarial machine learning can be regarded as two subjects participating in the game. They both aim to defeat the other party to strengthen their dominant position and continue to evolve. Moreover, their respective dominant positions vary according to different scenarios, but they are always opposed. For example, with regard to the maximum-minimization problem illustrated in Eq. (1), the adversarial attack is to minimize the reward, while the adversarial machine learning is to maximize the reward. However, in some classification tasks, it is a minimum-maximization problem. The adversarial attack is to maximize the loss function, while the adversarial machine learning is to minimize the loss function.

4 Challenges and opportunities for intelligent networking in adversarial environments

According to the discussion in Section 3, it is clear that adversarial environments can limit the development of AI technologies. However, if researchers actively solve these dilemmas, adversarial environments can further enhance the robustness and reliability. In this section, combining some representative scenarios in the field of intelligent networking and the essence revealed in Section 3, therefore, we will further analyze the situation of intelligent networking in adversarial environments, including challenges and opportunities.

Compared with computer vision and other fields, intelligent networking technologies start late. As of the current position, there are few studies [9, 10] that have focused on the intelligent networking in adversarial environments, which limits the further development of intelligent networking technologies. In this section, we analyze the challenges faced by several mainstream intelligent networking technologies in adversarial environments. Based on the opportunities that coexist with these challenges, we have proposed some extension frameworks and development approaches, hoping to pave the way for subsequent improvements in this area.

Regarding the specific scenarios of the intelligent networking, we first focus on the most representative RNN-based and DRL-based approaches, to illustrate the challenges and opportunities of intelligent networking technologies in adversarial environments. These two approaches support network security and network resource management, respectively. In addition to the internal technologies of the intelligent networking, we also pay attention to the collaborative learning based on networking. Collaborative learning can not only promote the development of intelligent networks [37], but also can be regarded as a network-based service.

4.1 RNN-based intelligent networking

Network traffic, as an important factor in the Internet, carries all Internet activities. In the face of various security threats in the network, traffic anomaly detection in cyberspace security, is an important research direction to defend against various network attacks. There have been some shallow learning methods to achieve traffic anomaly detection, but they are overly dependent on feature engineering, and it is difficult to deal with complex traffic anomalies, especially malicious intrusions that have not been seen before. Depending on a large number of neurons, DL can mine more effective features without the aid of feature engineering. At the same time, network traffic has obvious characteristics in sequence, so the RNN-based anomaly detection method [38, 39] has become one of the important intelligent networking technologies, which plays a vital role in proactive network management. For example, Yin et al. [38] proposed a method using RNN to implement an intrusion detection system (IDS), namely RNN-IDS. Experiments show that RNN-IDS outperforms other shallow learning methods in both binary and multi-classification tasks.

For network traffic, there are inevitable fluctuations. What's more, traffic monitoring tools are also subject to errors. These errors or fluctuations may become adversarial examples, causing the DNN-based model to lose its original function and triggering security incidents. Therefore, the traffic anomaly detection method based on DL must have strong robustness. Based on human knowledge, we can continuously optimize DNN architecture and improve the generalization ability [28]. For example, through theoretical knowledge such as information theory, Zhao et al. [28] proposed the stability-based defense mechanism to improve the generalization ability of intelligent edge computing services. Via the knowledge of frequency domain analysis, Fu et al. [15] proposed a realtime malicious traffic detection system, which achieves higher throughput, higher accuracy, and compatibility with more types of malicious traffic. However, due to the mystery of how neurons work, the improvement effect of this method is limited. In fact, we can actively use adversarial examples to improve the generalization ability of DNN, which is the opportunity brought by adversarial environments. Based on long short-term memory (LSTM), we for the first time propose an adversarial machine learning framework, illustrated in Figure 3. It can automatically optimize the network structure and improve the generalization ability. Specifically, it consists of LSTM-based generator and LSTM-based discriminator. In the newly proposed proactive defense framework[1)], the generator is an LSTM-based encoder-decoder structure. To obtain characteristics in sequence, massive real

1) Note that in this paper, the proactive defense framework we propose for RNN-based intelligent networking has universal characteristics and is compatible with other RNN variants (e.g., gate recurrent unit (GRU) and bi-directional LSTM (BiLSTM)) and scenarios similar to anomaly traffic detection. Therefore, we take LSTM as an example to illustrate the core ideas through a relatively intuitive description.

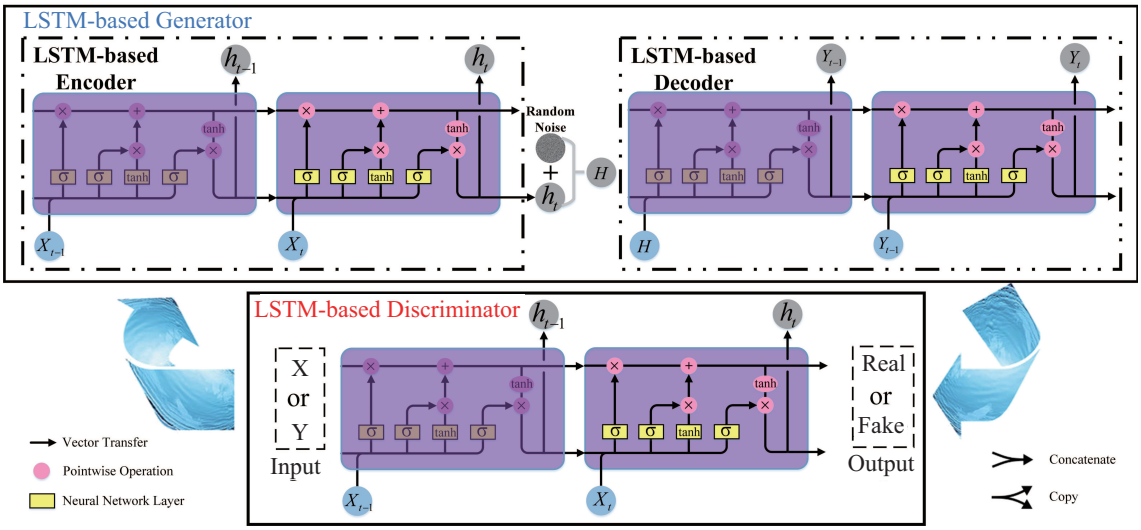

Figure 3 (Color online) LSTM-based adversarial machine learning framework for traffic anomaly detection, which can automatically optimize the DNN-based structure.

traffic can be fed to the LSTM-based encoder to optimize related model parameters, so that the referred LSTM-based encoder has the ability to characterize real traffic. After learning through several layers of networks, it can produce one hidden state, which contains both short-term and long-term memory. Subsequently, we actively add random noise to the hidden state, creating a new hidden state. And then, we use the new hidden state as input to the LSTM-based decoder, thereby reconstructing fake network traffic.

Based on the idea of game theory, we use another LSTM as a discriminator. Both real traffic and fake traffic generated by the generator, are then fed to the LSTM-based discriminator. Finally, the RNN-based proactive network management framework (e.g., anomaly traffic detection) will have a more reliable feature extraction capability, which is conducive to the construction of a more reliable and efficient intelligent networking system.

4.2 DRL-based intelligent networking

Owing to the characterization capabilities of DNN, reinforcement learning has achieved extraordinary success in the control areas, such as autonomous driving and edge network resource management. DRL does not require human intervention, but through continuous interactions between the DNN-based agent and the environment. Based on the feedback given by the environment, it can gradually find the best control strategies from scratch. This is consistent with the goal of intelligent control of networks. Therefore, in recent years, DRL has achieved rapid development and outstanding performance in the field of intelligent networking [3, 13, 40].

However, DRL, as a basic framework, has been found to face enormous challenges in adversarial environments [41, 42]. The core task of DRL is to use DNN to give optimal strategies in a given environment based on constant interactions with the environment. Some adversarial attacks against DRL have emerged recently. For example, the adversary can adopt the strategy-time attack method [41], which can reduce the rewards that the intelligent agent obtains from the environment by attacking the interaction for a short period of time. Specifically, in this small period of interaction, the adversary can create some unrealistic states that interfere with rewards. Moreover, Lin et al. [41] also proposed another effective attack strategy for DRL, namely the enchanting attack. It predicts the future states, and then generates a preferred sequence of actions for luring the agent. Subsequently, a sequence of adversarial examples is crafted to lure the agent to learn wrong strategies.

Similar attacks are also easy to implement in the field of intelligent networking, but few people are currently concerned. As illustrated in Figure 4, in the field of intelligent networking, the states contained in the environment include but not limited to the basic 5-tuple (i.e., protocol, source IP, destination IP, source port, and destination port), throughput, and round-trip time (RTT). For these state elements, it is very easy for hacker or adversary to interfere with, or design attack samples based on predictions. Note that the attack here only requires the attacker to have the opportunity to interact with the victim [42],

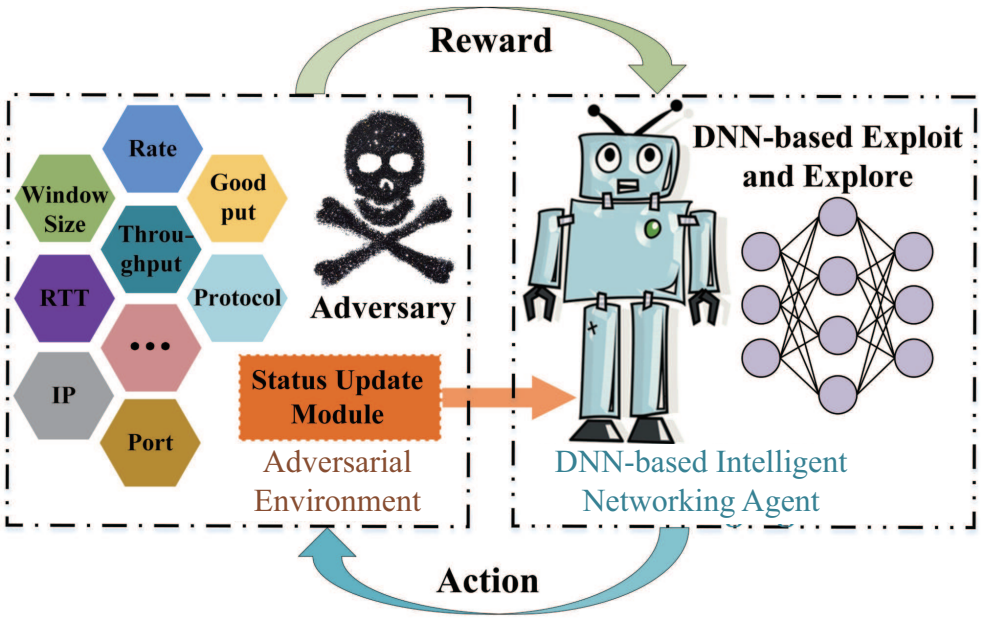

Figure 4 (Color online) Examples of DRL-based intelligent network resource management in the adversarial environment.

and does not require the attacker to have control authority over the victim as a prerequisite for the attack. DRL-based intelligent networking technologies have to learn the optimal strategy based on the feedback from the environment. Once the hacker successfully modifies the state based on his or her own intention, the reward calculated based on the feedback from the environment will be wrong. And then, it is easy to learn some wrong strategies, even a strategy designed by the hacker. Considering the widespread utilization of the intelligent networking, similar attacks can easily cause irreversible consequences.

To defend against similar attacks for DRL-based intelligent networking, we can characterize networking adversarial examples based on specific properties of the network environment. For example, temporal dependency in network traffic can be used to gain discriminative capacity against adversarial examples. In addition to characterizing adversarial examples, we can leverage these attacks against DRL to train a more robust DRL-based intelligent networking framework. Through the adversarial training, intelligent networking technologies based on DRL can get significant increase in robustness to parameter variations. It can enable intelligent networking technologies to provide more reliable services, thus promoting the further popularization and deployment of intelligent networks.

Overall, the game theoretic framework for adversarial attack and defense of intelligent networking in adversarial environment can be described as the maximum-minimization problem. In this paper, we utilize

$$\max_{\boldsymbol{\theta}} \mathbb{E}_{(s_t,a_t)\sim\mathcal{D}}\left[\min_{\delta_t\in\Phi(s_t)} \mathcal{R}(s_t+\delta_t, a_t; \boldsymbol{\theta})\right] \tag{1}$$

to illustrate the game theoretic framework.

Regarding Eq. (1), s_t and a_t are the environment state and the action performed by the DNN-based intelligent networking agent at time t, respectively. $\mathcal{D}$ refers to the distribution of related data. $\mathcal{R}(\cdot)$ refers to the rewards obtained by performing the specific action with the specific environment state. Through continuous interactive training, the DNN-based intelligent networking agent aims to find the optimal model parameters $\boldsymbol{\theta}$ to maximize the expected cumulative reward $\mathbb{E}$. On the contrary, the attacker relies on limited knowledge (e.g., model structure and observed data laws) to create perturbation δ_t with specific constraint $\Phi(\cdot)$. In other words, $s_t+\delta_t$ is the adversarial example. The referred constraint $\Phi(s_t)$ here is to ensure the concealment of the attack, so $s_t+\delta_t$ is a crafted adversarial example. The attacker aims to use adversarial examples to minimize the reward, thereby forcing the DNN-based intelligent networking agent to learn some wrong strategies.

4.3 Collaborative learning for intelligent networking

With the rapid development of network technologies, the network is not only integrated into all aspects of human life, but also connected to multiple smart devices. In particular, the rise of autonomous 5G

has further promoted the development of IoT devices. For autonomous 5G, IoT-related automated 5G technologies are much more important. IoT devices or IoT-like devices, computing power, and storage capacity are very limited. However, the success of DL algorithms is mainly due to the computing power and the access to significant amounts of data. In the process of accessing more data, the issue of privacy leakage in data sharing cannot be ignored. For example, Qu et al. [43] proposed a GAN-based data argumentation method to cloak original data information, thereby protecting privacy in data sharing. Data sharing based on privacy protection can enrich the data of a single device, but the limited computing ability of IoT devices is still a bottleneck.

To enable various resource-constrained devices, like IoT devices, to enjoy the benefits of DL, collaborative learning for indirect data sharing has become one of the important ways [28, 37]. In this way, collaborators can independently learn based on their own small amount of data and limited computing power, then obtain a comprehensive DL-based intelligent model through cooperation between collaborators. Moreover, the Internet architecture is naturally distributed. A large number of devices (e.g., router and switch) are connected to each other and cooperate with each other to perform various tasks such as network transmission. Collaborative learning is perfectly adapted to this scenario, thereby promoting the development of intelligent networking.

With the development of adversarial attacks, privacy leakage has brought enormous challenges to collaborative learning. The communication-efficiency problem can affect the quality of service [44], while the privacy leakage may directly cause users to refuse to adopt the service. Specifically, in collaborative learning, each collaborator only trains their respective network structures in order to maintain their own privacy data. Subsequently, the collaborators upload model parameters they have learned to the cloud and then synchronize them to other collaborators, or directly to other collaborators through the peer-to-peer network, so that each collaborator can learn a relatively complete model. However, even with technologies such as differential privacy to protect information, it is still susceptible to adversarial attacks. For example, the adversary can pretend to be a collaborator while training a generated adversarial network locally. It can guess what privacy data other honest collaborators have by generating some fake samples [45]. This type of attack not only breaks the defense of the differential privacy, but is also easier to implement than attacking cloud-based centralized learning. Since collaborative learning requires the participation of multiple collaborators, and the impact of backdoor attacks against DL-based models is more secretive, the malicious behavior of backdoors can also be propagated from one collaborator to other collaborators [28]. Especially for privacy-sensitive applications and scenarios, similar attacks to collaborative learning are extremely serious.

Similar to the DL-based model inherent in intelligent networking, attacks to collaborative learning also provide the opportunity to proactively design a more resilient and reliable collaborative learning mechanism for intelligent networking. For example, to defend against privacy leaks based on adversarial attacks, we can apply blockchain to ensuring privacy protection during parameter transmission. Meanwhile, blockchain-based storage technology can also detect malicious activities such as data tampering involved in backdoor attacks. In addition, we can design a more reliable differential privacy mechanism to enable the privacy-preserving collaborative learning.

5 Conclusion and future work

Since more and more intelligent networking technologies are implemented with DL, while DNN-based models are vulnerable to crafted adversarial examples, understanding existing intelligent networking technologies in adversarial environments is becoming significant for safety-critical or robustness-critical scenarios. In this paper, we present a comprehensive analysis of opportunities and challenges for intelligent networking in adversarial environments, including adversarial attacks and adversarial machine learning. Specifically, in terms of RNNs and DRL, both of which are widely used frameworks for intelligent networking, we analyze the implementation of attacks and propose solutions. In addition, collaborative learning, which is inherently compatible with the distributed Internet, is analyzed in detail. To the best of our knowledge, this is the first paper that provides an extensive analysis of existing intelligent networking technologies in adversarial environments and proposes valuable frameworks and approaches to pave the way for future advancements.

In addition to attracting widespread attention from academia and the industry in the field of intelligent networking through this paper, we think that there are still some open research issues to achieve an

extensive deployment of intelligent networking in adversarial environments. Specifically, we emphasize the following open research issues to further pave the way for the widespread deployment of intelligent networking in adversarial environments.

(1) Simulation platform and dataset for adversarial attacks. The vulnerability of intelligent networking will cause significant economic losses or security accidents. We cannot directly perform adversarial attacks against intelligent networking. Therefore, in-depth research requires researchers to design simulation platforms that can create various attack scenarios, and then design related attack methods, as well as defense mechanisms. In addition, how to integrate or expand historical attack data from different scenarios into a dataset is extremely important for promoting the development of intelligent networking in adversarial environments.

(2) Robustness evaluation standard. Whether an attack method is efficient or a defense mechanism is effective is inseparable from the robustness evaluation of the relevant intelligent networking technology. However, intelligent networking technologies involve diverse scenarios, and adversarial environments are complex and changeable. To ensure the safety and reliability of intelligent networking technology, researchers are required to design robustness evaluation standards with high flexibility and strong scalability.

Different from the above research directions that are applicable to various issues of intelligent networking technologies, some specific academic issues are also worthy of attention. As discussed in Subsection 4.3, the interconnected nature of network devices makes collaborative learning an important path for the development of intelligent networking. However, multi-party cooperation makes malicious attacks concealed. In addition to the simulation platform and evaluation standard, we must also actively defend against malicious adversarial attacks from the design of collaborative algorithms.

Acknowledgements This work was in part supported by National Science Foundation for Distinguished Young Scholars of China (Grant No. 61825204), National Natural Science Foundation of China (Grant Nos. 61932016, 62132011), Beijing Outstanding Young Scientist Program (Grant No. BJJWZYJH01201910003011), China Postdoctoral Science Foundation (Grant No. 2021M701894), and China National Postdoctoral Program for Innovative Talents. Dan WANG's work is supported in part by General Research Fund (Grant Nos. 15210119, 15209220, 15200321), Innovation Technology Fund (ITSP Program ITS/070/19FP), Collaborative Research Fund (Grant Nos. C5026-18G, C5018-20G), The Hong Kong Polytechnic University (Grant No. 1-ZVPZ), and a Huawei Collaborative Project. We also thank anonymous reviewers for their comments and guidance.

References

1 Cabaj K, Mazurczyk W, Nowakowski P, et al. Towards distributed network covert channels detection using data mining-based approach. In: Proceedings of the 13th International Conference on Availability, Reliability and Security, 2018. 12
2 Mirsky Y, Doitshman T, Elovici Y, et al. KitSune: an ensemble of autoencoders for online network intrusion detection. In: Proceedings of Network and Distributed Systems Security Symposium, 2018
3 Chen L, Lingys J, Chen K, et al. AuTO: scaling deep reinforcement learning for datacenter-scale automatic traffic optimization. In: Proceedings of ACM SIGCOMM, 2018. 191–205
4 Bega D, Gramaglia M, Fiore M, et al. AZTEC: anticipatory capacity allocation for zero-touch network slicing. In: Proceedings of IEEE INFOCOM, 2020. 794–803
5 Zhao Y, Qiao M N, Wang H Y, et al. TDFI: two-stage deep learning framework for friendship inference via multi-source information. In: Proceedings of IEEE INFOCOM, 2019. 1981–1989
6 Benzaid C, Taleb T. AI-driven zero touch network and service management in 5G and beyond: challenges and research directions. IEEE Network, 2020, 34: 186–194
7 Lei K, Liang Y Z, Li W. Congestion control in SDN-based networks via multi-task deep reinforcement learning. IEEE Network, 2020, 34: 28–34
8 Gong S M, Lu X, Hoang D T, et al. Toward smart wireless communications via intelligent reflecting surfaces: a contemporary survey. IEEE Commun Surv Tut, 2020, 22: 2283–2314
9 Lin Y, Zhao H J, Tu Y, et al. Threats of adversarial attacks in DNN-based modulation recognition. In: Proceedings of IEEE INFOCOM, 2020. 2469–2478
10 Sagduyu Y E, Shi Y, Erpek T. Adversarial deep learning for over-the-air spectrum poisoning attacks. IEEE Trans Mobile Comput, 2021, 20: 306–319
11 Qiu H, Dong T, Zhang T W, et al. Adversarial attacks against network intrusion detection in IoT systems. IEEE Internet Things J, 2021, 8: 10327–10335
12 Han D Q, Wang Z L, Zhong Y, et al. Evaluating and improving adversarial robustness of machine learning-based network intrusion detectors. IEEE J Sel Areas Commun, 2021, 39: 2632–2647
13 Xu Z Y, Tang J, Yin C X, et al. Experience-driven congestion control: when multi-path TCP meets deep reinforcement learning. IEEE J Sel Areas Commun, 2019, 37: 1325–1336
14 Boyan J A, Littman M L. Packet routing in dynamically changing networks: a reinforcement learning approach. In: Proceedings of Conference and Workshop on Neural Information Processing Systems, 1994. 671–678
15 Fu C P, Li Q, Shen M, et al. Realtime robust malicious traffic detection via frequency domain analysis. In: Proceedings of ACM SIGSAC Conference on Computer and Communications Security, 2021. 3431–3446
16 Liu C Y, Xu M W, Yang Y, et al. DRL-OR: deep reinforcement learning-based online routing for multi-type service requirements. In: Proceedings of IEEE INFOCOM, 2021
17 Yan S Y, Wang X L, Zheng X L, et al. ACC: automatic ECN tuning for high-speed datacenter networks. In: Proceedings of ACM SIGCOMM, 2021. 384–397

18 Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Proceedings of Conference and Workshop on Neural Information Processing Systems, 2014. 2672–2680
19 Ma X J, Li B, Wang Y S, et al. Characterizing adversarial subspaces using local intrinsic dimensionality. In: Proceedings of International Conference on Learning Representations, 2018
20 Li J, Liu Y, Chen T, et al. Adversarial attacks and defenses on cyber-physical systems: a survey. IEEE Internet Things J, 2020, 7: 5103–5115
21 Wang N, Chen Y M, Hu Y, et al. MANDA: on adversarial example detection for network intrusion detection system. IEEE Trans Depend Secure Comput, 2022. doi: 10.1109/TDSC.2022.3148990
22 Treu M, Le T N, Nguyen H H, et al. Fashion-guided adversarial attack on person segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2021. 943–952
23 Chen Y X, Yuan X J, Zhang J S, et al. Devil's whisper: a general approach for physical adversarial attacks against commercial black-box speech recognition devices. In: Proceedings of USENIX Security, 2020. 2667–2684
24 Wu F, Long Y H, Zhang C, et al. LinkTeller: recovering private edges from graph neural networks via influence analysis. In: Proceedings of IEEE Symposium on Security and Privacy (SP), 2022
25 Xie C L, Chen M H, Chen P Y, et al. CRFL: certifiably robust federated learning against backdoor attacks. In: Proceedings of International Conference on Machine Learning, 2021. 11372–11382
26 Yatsura M, Metzen J, Hein M. Meta-learning the search distribution of black-box random search based adversarial attacks. In: Proceedings of Conference and Workshop on Neural Information Processing Systems, 2021
27 Chivukula A S, Liu W. Adversarial deep learning models with multiple adversaries. IEEE Trans Knowl Data Eng, 2019, 31: 1066–1079
28 Zhao Y, Xu K, Wang H Y, et al. Stability-based analysis and defense against backdoor attacks on edge computing services. IEEE Network, 2021, 35: 163–169
29 Hameed M Z, Gyorgy A, Gunduz D. The best defense is a good offense: adversarial attacks to avoid modulation detection. IEEE Trans Inform Forensic Secur, 2020, 16: 1074–1087
30 Usama M, Mitra R, Ilahi I, et al. Examining machine learning for 5G and beyond through an adversarial lens. IEEE Internet Comput, 2021, 25: 26–34
31 Zanella-Beguelin S, Tople S, Paverd A, et al. Grey-box extraction of natural language models. In: Proceedings of International Conference on Machine Learning, 2021. 12278–12286
32 Goodfellow I J, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. In: Proceedings of International Conference on Learning Representations, 2015
33 Madry A, Makelov A, Schmidt L, et al. Towards deep learning models resistant to adversarial attacks. In: Proceedings of International Conference on Learning Representations, 2018
34 Moosavi-Dezfooli S M, Fawzi A, Frossard P. Deepfool: a simple and accurate method to fool deep neural networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016. 2574–2582
35 Carlini N, Wagner D. Towards evaluating the robustness of neural networks. In: Proceedings of IEEE Symposium on Security and Privacy (SP), 2017. 39–57
36 Yang Z L, Li B, Chen P Y, et al. Characterizing audio adversarial examples using temporal dependency. In: Proceedings of International Conference on Learning Representations, 2019
37 Wang X F, Han Y W, Wang C Y, et al. In-Edge AI: intelligentizing mobile edge computing, caching and communication by federated learning. IEEE Network, 2019, 33: 156–165
38 Yin C L, Zhu Y F, Fei J L, et al. A deep learning approach for intrusion detection using recurrent neural networks. IEEE Access, 2017, 5: 21954–21961
39 Diro A, Chilamkurti N. Leveraging LSTM networks for attack detection in fog-to-things communications. IEEE Commun Mag, 2018, 56: 124–130
40 Liang E, Zhu H, Jin X, et al. Neural packet classification. In: Proceedings of ACM SIGCOMM, 2019. 256–269
41 Lin Y C, Hong Z W, Liao Y H, et al. Tactics of adversarial attack on deep reinforcement learning agents. In: Proceedings of International Joint Conference on Artificial Intelligence, 2017. 3756–3762
42 Wang F, Zhong C, Gursoy M C, et al. Defense strategies against adversarial jamming attacks via deep reinforcement learning. In: Proceedings of the 54th Annual Conference on Information Sciences and Systems (CISS), 2020. 1–6
43 Qu Y Y, Zhang J W, Li R D, et al. Generative adversarial networks enhanced location privacy in 5G networks. Sci China Inf Sci, 2020, 63: 220303
44 Liu Y, Zhao Y, Zhou G M, et al. FedPrune: personalized and communication-efficient federated learning on non-IID data. In: Proceedings of International Conference on Neural Information Processing, 2021. 430–437
45 Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: information leakage from collaborative deep learning. In: Proceedings of ACM SIGSAC Conference on Computer and Communications Security, 2017. 603–618