Momentum-Based Adversarial Attacks Against End-to-End Communication Systems

Qiuna Zhang, Yongkui Ma, Honglin Zhao(✉), Chengzhao Shan, and Jiayan Zhang

Communication Research Center, Harbin Institute of Technology, Harbin, China
{yk_ma,hlzhao,czshan,jyzhang}@hit.edu.cn

Abstract. Deep learning (DL) based communication system is a promising novel architecture to implement end-to-end optimization compared with conventional block-separated optimization schemes. However, the vulnerability to adversarial examples of deep neural networks poses significant security concern on the end-to-end communication systems. Adversarial attacks serve as a fundamental surrogate to evaluate the robustness of the DL-based communication systems before they are deployed. Specifically, we propose a new adversarial attack method with momentum iterative gradient against the end-to-end communication systems. For targeted attacks, embedding the momentum term in the iterative process can help loss function stabilize the update direction and avoid getting stuck in saddle points and poor local minima. Therefore, the momentum-based method can enhance the effectiveness without losing the transferability of adversarial attacks. Numerous simulation results illustrate that the proposed method can achieve superior block error rate compared with traditional jamming attacks and no momentum accumulated adversarial attacks.

Keywords: Momentum · Adversarial attacks · End-to-end communication systems · Deep learning · Wireless security · Model robustness

1 Introduction

Due to powerful nonlinear approximation and optimization capabilities, deep learning (DL) has become a promising technique to satisfy the growing demands of the fifth-generation (5G) wireless communications and beyond, such as high reliability, ultra-high capacity and low-latency. Recently, abundant concepts and applications of deep learning have been deployed in the field of communications. One of the most novel concepts is an end-to-end communication system based on an autoencoder framework [1], and a series of studies have been conducted around it [2–5].

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 520–532, 2022.
https://doi.org/10.1007/978-3-030-93398-2_48

Whereas, deep neural networks (DNNs) are exceedingly susceptible to adversarial examples [6,7], these examples can fool classifiers by adding small and human-imperceptible perturbations to legitimate examples. This unique characteristic of DNN raises significant security and robustness concern about implying DL method in the field of communications, especially for the end-to-end communication system based on the autoencoder architecture, whose transmitter, channel and receiver are composed of DNNs. Compared with conventional jamming attacks, adversarial attacks are more destructive and unnoticeable, because adversarial perturbations are essentially optimization vectors deliberately crafted in feature space, which can change the optimization of loss gradients in wrong directions. For deep learning based communication systems, owing to the openness of wireless communication channel, an illegitimate attacker can add a small perturbation to the transmitted signal when it passes through the channel, thus confuses the receiver [8]. As a consequence, how to improve the robustness of the DL-based communication systems is an urgent challenge. Therefore, for the DL-based communication systems, it is necessary to conduct further research from the perspective of adversarial attacks, since it inspires the study of defense methods, as well as facilitates the robustness assessment.

For the DL-based communication systems, despite broad studies on various aspects, poor discussions of adversarial attacks are conducted around them. In [8] and [9], one-step gradient based perturbation generating algorithm is adopted to craft white-box and black-box attacks. More transferable adversarial examples can be generated by one-step gradient based attack methods, however, they usually have a low success rate for attacking white-box models [10]. Therefore, we consider applying momentum into adversarial attacks against DL-based communication systems, aiming to improve the effectiveness of crafting white-box attacks without reducing their transferability.

In this paper, we investigate momentum-based adversarial attacks against end-to-end communication systems. The main contributions of our work are: (i) We propose a momentum-based adversarial perturbation generation algorithm to generate adversarial perturbations for end-to-end communication systems, in which we utilize velocity vector to stabilize the optimization direction of loss gradient, so as to escape from local minima and saddle points. (ii) Through crafting targeted attacks against autoencoders with different structures, we verify that momentum-based adversarial attacks can achieve better block error rate (BLER) than traditional jamming attacks, which demonstrates their powerful destructiveness. (iii) By crafting white-box and black-box attacks, we show that adversarial attacks with momentum are more destructive than attacks without momentum accumulation in [8], which reveals that momentum-based attacks can increase the attack effectiveness of white-box models without losing the transferability of attacking black-box models.

2 System Model

We consider a DL-based communication system consisting of transmitter, channel and receiver, which can be represented as an autoencoder since its optimization object is to reconstruct its input at the output [1]. The characteristic of the

DL-based communication system is that all its transmitter, channel and receiver are composed of deep neural networks, in which both transmitter and receiver are trainable. Nevertheless, in most structures, the channel layer is untrainable, unless the channel state information (CSI) is unknown, in this setting, a channel modeling method based on generative adversarial network (GAN) can be used to train the channel layer [3].

The DL-based communication system tries to learn the channel characteristic (noise jamming, fading and distortion) mapped from the transmitter to the receiver, so that the massage sent by the transmitter can be recovered at the receiver side with minimum error rate. Corresponding to the components of the autoencoder, the transmitter and receiver are represented by the encoder and decoder, respectively.

In the end-to-end communication system, k bit message $s \in \{1, 2, \ldots, M\}$ first passes through the transmitter (encoder) to the channel, where it may be attacked by illegitimate attackers using adversarial attack methods, and then the receiver (decoder) tries to recover it. Before input to the encoder, s needs to be represented as a one-hot vector of dimension $M = 2^k$, and then the encoder transforms it by applying $f : \mathbb{R}^M \rightarrow \mathbb{R}^{2n}$ to generate the transmitted signal $\mathbf{x} = f(s) \in \mathbb{R}^{2n}$ for n complex channel uses [2]. However, in practice, $2n$ real channel uses are usually used to replace the n complex channel uses, since there is no complex operation in the actual operation of DNN [1]. The last layer of the encoder is a normalization layer constrained by $\mathbb{E}\left[x_i^2\right] \leq 0.5, \ \forall i.$, where $\mathbb{E}[\cdot]$ is the expectation of the elementwise square of $\mathbf{x}$, which ensures the power constraint before $\mathbf{x}$ is transmitted to the channel [1].

When $\mathbf{x}$ is transmitted to the wireless channel, it may face various channel environments. In this work, as set in [8], we consider the additive white Gaussian noise (AWGN) channel realized by a noise layer, in which illegitimate attackers could craft adversarial perturbations to attack the transmitted signal $\mathbf{x}$. Therefore, the output of the channel, namely, the received signal of the receiver $\mathbf{y}$ is given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n} + \mathbf{p} \tag{1}$$

where $\mathbf{n}$ is a Gaussian distribution vector and $\mathbf{p}$ is an adversarial perturbation. The noise vector $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$, whose variance $\sigma^2 = (2RE_b/N_0)^{-1}$. For σ^2, R represents the data rate, E_b denotes the energy per bit, and N_0 is the noise power spectral density. As for the adversarial perturbation $\mathbf{p}$, it is delicately crafted for the characteristic of DNN by illegitimate attackers.

After the channel, the receiver (decoder) tries to achieve the transformation $g : \mathbb{R}^{2n} \rightarrow \mathbb{R}^M$ for the received signal $\mathbf{y}$, to generate a reconstructed message $\hat{s}$. The last layer of the decoder realizes softmax activation operation [11], and the estimated message $\hat{s}$ is the index of the highest probability element in the M dimensional probability vector b, which is the output of the softmax activation function. The composition of the end-to-end communication system under adversarial attacks is shown in Fig. 1.

To express the reconstruction problem of communication message as a classification mission, the cross-entropy loss function is adopted to optimize the distance between s and b. Thus, the cross-entropy loss function L is written as

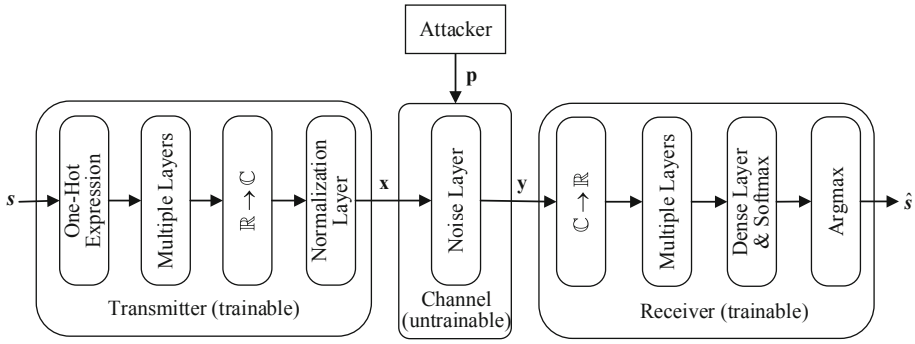

Fig. 1. An end-to-end communication system represented as an autoencoder under an adversarial attack.

$$L = -\sum_{i} s_i \log(b_i) \tag{2}$$

where s_i and b_i are corresponding ith element of s and b, respectively.

Generally, block error rate (BLER) P_e is used to measure the performance of the DL-based communication systems, defined as

$$P_e = \frac{1}{M} \sum_{s} \Pr(\hat{s} \neq s|s) \tag{3}$$

It is worth noting that the BLER in the end-to-end communication systems is equivalent to the symbol error rate (SER) in conventional communication systems.

3 A Momentum-Based Adversarial Perturbation Generation Algorithm

The concept of momentum was first proposed in [12], as a tool to help improve gradient descent algorithms by accumulating the gradients of loss function in previous iterations as a velocity vector. [13] applies the concept of momentum to generate adversarial examples for image adversarial attacks, inspired by this, we incorporate momentum into adversarial perturbation generation of the end-to-end communication systems. [14] shows that the accumulation of previous gradients could help DNN to avoid drawbacks of local minimization, so in targeted attacks, we can leverage momentum to help the current gradient to barrel through the critical points of loss surface, including saddle points and poor local minima or maxima. Moreover, from [15] we konw that the momentum-based gradient iteration method could also achieve better stability for the update process of stochastic gradient descent, which plays an important role in the effectiveness of adversarial attacks. Therefore, based on the analysis mentioned above, we propose a momentum-based adversarial perturbation generation method, which is formulated in this section.

3.1 Momentum-Based Iterative Gradient Method

Compared with untargeted adversarial attacks, targeted adversarial attacks are easier to implement and more efficient, so we focus on targeted attacks. In order to generate a targeted adversarial example $\mathbf{x}^{adv}$ based on a real example $\mathbf{x}$, it should be constrained by L_p norm within accuracy constraint ε, to make sure the adversarial attack is unnoticeable. Therefore, the gradient-based methods solving the constrained optimization problem as below to seek the optimal adversarial example

$$\begin{aligned} &\arg\min_{\mathbf{x}^{adv}} J(\mathbf{x}^{adv}, y^{target}) \\ &\text{s.t.} \left\|\mathbf{x}^{adv} - \mathbf{x}\right\|_p \leq \varepsilon \end{aligned} \tag{4}$$

The gradient iteration methods can be easily generalized to the attack setting of explicit norm bound constrains, such as L_1, L_2, L_∞ norm bounds. Using the cumulative gradient of all previous steps to substitute current gradient, we can extend any gradient iteration approach to its momentum gradient variant. However, in the field of wireless communications, compared to other norm bounds, L_2 norm seems to be a more appropriate choice since it usually represents the power of the adversarial perturbation [9]. Therefore, we only introduce targeted attack methods for generating adversarial perturbations in terms of L_2 norm bound.

For targeted attacks, to seek an adversarial perturbation within the vicinity of a real example subject to L_2 distance, $\left\|\mathbf{x}^{adv} - \mathbf{x}\right\|_2 \leq \varepsilon$, the update process of the momentum-based gradient method incorporates accumulated gradient of all previous steps into current gradient, which can be written as

$$\mathbf{g}_t = \mu \cdot \mathbf{g}_{t-1} + \alpha \cdot \frac{\nabla J(\mathbf{x}_t^{adv}, y^{target})}{\left\|\nabla J(\mathbf{x}_t^{adv}, y^{target})\right\|_2} \tag{5}$$

where μ is the decay factor, $\mathbf{g}_{t-1}$ denotes the accumulated gradient up to the current iteration, y^{target} is defined as the targeted misclassification label, α is the stepsize.

In consequence, the targeted momentum-based iterative gradient within a L_2 norm constraint is

$$\mathbf{x}_t^{adv} = \mu \cdot \mathbf{x}_{t-1}^{adv} - \alpha \cdot \frac{\mathbf{g}_t}{\|\mathbf{g}_t\|_2} \tag{6}$$

3.2 Generative Algorithm of Momentum-Based Adversarial Perturbation

In order to evaluate the robustness of the wireless communication systems based on DL, we propose a momentum-based adversarial attack method against the end-to-end communication systems. Adversarial attack method with momentum generates adversarial perturbations, which can fool the receiver (decoder), resulting in superior increasing of BLER. Algorithm 1 presents how to generate an adversarial perturbation based on the momentum iterative gradient. In Algorithm 1, we use bisection search to find an appropriate value of the stepsize α

among all possible targeted categories as in [8]. Furthermore, for a specific target class, the momentum-based algorithm tries to minimize $J(\mathbf{x}_t^*, y^{target})$, and then use the bisection search to find appropriate gradient after T iterations with momentum.

Algorithm 1. Generating a Momentum-based Adversarial Perturbation

Input: A real input $\mathbf{x}$ and its ground-truth label y_{true}, the pretrained model f with loss function J, desired perturbation accuracy ε_{acc}, the number of iterations T, maximum allowed perturbation norm p_{max} and decay factor μ.

Output: An adversarial perturbation $\mathbf{p}_{\mathbf{x}}^{adv}$ of input $\mathbf{x}$, an adversarial example $\mathbf{x}^{adv}$ with constraint $\left\|\mathbf{x}^{adv} - \mathbf{x}\right\|_2 \leq \varepsilon$.

Initialization: Set the initial value $\mathbf{g}_0 = 0$ and $\mathbf{x}_0^* = \mathbf{x}$.
for *class-index* in *range*(C) **do**
  $\varepsilon_{\max} \leftarrow p_{\max}$, $\varepsilon_{\min} \leftarrow 0$, $\alpha \leftarrow \varepsilon_{\max} + \varepsilon_{\min}/2$
  Input $\mathbf{x}_t^*$ to pretrained model f for obtaining the gradient $\nabla_{\mathbf{x}} J(\mathbf{x}_t^*, y^{target})$
  for $t = 1$ to T **do**
    Update $\mathbf{g}_t$ by accumulating the velocity vector in the gradient direction as
    $\mathbf{g}_t = \mu \cdot \mathbf{g}_{t-1} + \alpha \cdot \nabla_{\mathbf{x}} J(\mathbf{x}_t^*, y^{target})(\left\|\nabla_x J(\mathbf{x}_t^*, y^{target})\right\|_2)^{-1}$
    while $\varepsilon_{\max} - \varepsilon_{\min} > \varepsilon_{acc}$ **do**
      $\varepsilon_{ave} = \varepsilon_{\max} + \varepsilon_{\min}/2$ and $\alpha = \varepsilon_{ave}$
      Update $\mathbf{x}_t^*$ by applying the momentum gradient as
      $\mathbf{x}_t^* = \mathbf{x}_{t-1}^* - \alpha \cdot \mathbf{g}_t(\|\mathbf{g}_t\|_2)^{-1}$
      if $f(\mathbf{x}_t^*) = y_{true}$ **then**
        $\varepsilon_{\min} \leftarrow \varepsilon_{ave}$
      else
        $\varepsilon_{\max} \leftarrow \varepsilon_{ave}$
      end if
    end while
    $\varepsilon_{class-index} = \varepsilon_{\max}$
  end for
  $\mathbf{x}_{class-index}^* = \mathbf{x}_T^*$
end for
$target - class_{\varepsilon} = \operatorname{argmin} \varepsilon_{class-index}$ and $\alpha^* = \min \varepsilon_{class-index}$
$target - class_{\mathbf{g}} = \operatorname{argmin} \mathbf{g}_{class-index}$ and $\mathbf{g}_T^* = \min \mathbf{g}_{class-index}$
return $\mathbf{x}^{adv} = \mathbf{x}_{target-class_{\mathbf{g}}}^*$, $\mathbf{p}_{\mathbf{x}}^{adv} = \alpha^* \cdot \mathbf{g}_T^*(\|\mathbf{g}_T^*\|_2)^{-1}$

In our experiments, we only report the results of targeted attacks using momentum-based perturbation generation algorithm. We set the decay factor μ to 1.0 and the number of iterations T to 20. The decay factor μ plays an instrumental role in increasing the BLER of momentum-based algorithms. In order to find the appropriate value of the decay factor, we attack the end-to-end communication systems based on CNN and MLP respectively with PSR $= -6$ dB, $E_b/N_0 = 8$ dB, while the decay factor ranging from 0.0 to 2.0 with a granularity 0.1. In Fig. 2, the curve of BLER for end-to-end communications with different values of μ does not change much, but it can achieve stable performance when $\mu = 1.0$ for different network structures, so we choose $\mu = 1.0$ as one of our

hyper-parameters, just like in [13], which means we simply add up all gradients of previous iterations to update the current gradient.

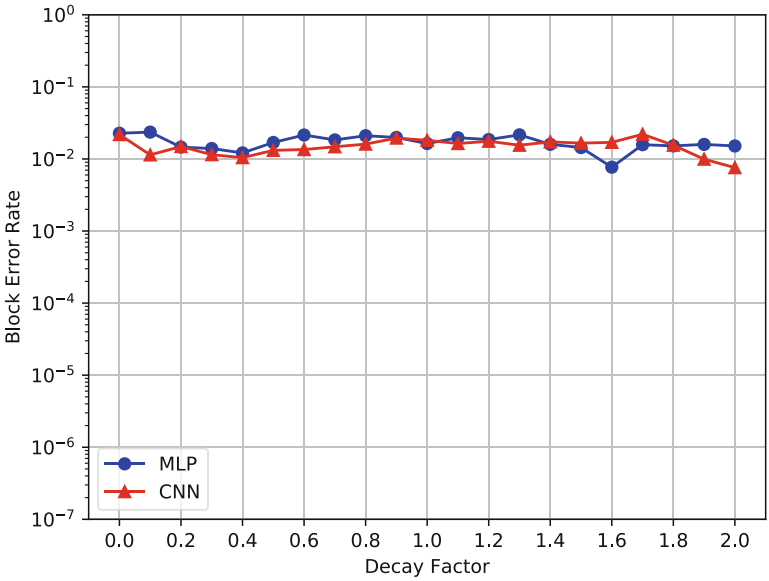

Fig. 2. BLER versus the decay factor μ under the momentum-based attacks against the MLP autoencoder and CNN autoencoder.

We also investigate the effect of the number of iterations T on BLER when using the momentum-based adversarial attacks. We adopt the hyper-parameter $\mu = 1$ to attack end-to-end communication systems with T ranging from 1 to 20. We evaluate the BLER of adversarial perturbations against autoencoders in Fig. 3. It can be observed that the BLER of momentum-based algorithm against autoencoders with different structures maintains a stable value. It proves that the adversarial perturbations generated by momentum iterative methods are difficult to overfit a white-box model and maintain stable effectiveness for different models. However, from the result of simulation, it is obvious that when $T = 20$, the BLER of different network structures are better than others, so we choose $T = 20$ as the value of another hyper-parameter.

4 Crafting Momentum-Based Adversarial Attacks Against End-to-End Communication Systems

Equation (1) shows how an adversarial perturbation is added to the transmitted signal in the wireless channel, therefore, the optimization objective of

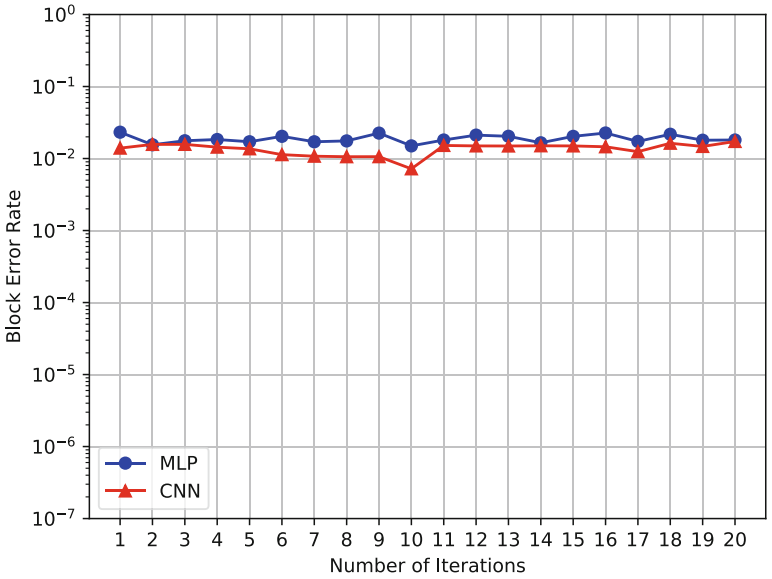

Fig. 3. BLER versus the number of iterations T under the momentum-based attacks against the MLP autoencoder and CNN autoencoder.

the attacker is to craft an adversarial perturbation to confuse the considered receiver(decoder) as in [8]

$$\begin{aligned} &\min_{\mathbf{p}} \|\mathbf{p}\|_2 \\ &\text{s.t. } g(\mathbf{x}+\mathbf{n}+\mathbf{p}) \neq g(\mathbf{x}+\mathbf{n}) \end{aligned} \tag{7}$$

From [6], we know that the Eq. (7) does not belong to convex function because the receiver mapping g does not have a convex structure. [8] uses fast gradient method without momentum [6,7] to approximate the optimum solution of Eq. (7), which would face the dilemma of being trapped by poor local minima or saddle points. By using the momentum-based gradient method, we expect to provide a velocity vector in the gradient direction of the loss surface across iterations, which can feed the demand of stabilizing the update direction as well as escaping from poor local minima and saddle points.

The structures of autoencoder generally used as the benchmarks in end-to-end communication systems are multi-layer perceptron (MLP) and convolutional neural network (CNN). To enable a performance comparison for the adversarial attacks against the end-to-end communication systems without momentum, we adopt the same autoencoder structures and dimensions as in [8], namely a MLP-based autoencoder and a CNN-based autoencoder which are shown in Table 1 and Table 2 respectively. The results of BLER are obtained by Monte Carlo simulation, and the number of simulation is 1,000,000. All the simulations are performed using TensorFlow on an NVIDIA GeForce GTX 1080 Ti graphic processing unit.

Table 1. The structure and dimension of the considered autoencoder based on MLP.

Block structure	Layer	Output dimension
Encoder	One-Hot Input	M
	Dense + ReLU	M
	Dense + Linear	2n
	Power Normalization	2n
Channel	AWGN Layer (+Adversarial Attack)	2n
Decoder	Dense + ReLU	M
	Dense + Softmax	M

Table 2. The structure and dimension of the considered autoencoder based on CNN.

Block structure	Layer	Output dimension
Encoder	One-Hot Input	M
	Dense + eLU	M
	Conv1D + eLU	$16 \times$ M
	Conv1D + eLU	2n
	Power Normalization	2n
Channel	AWGN Layer (+Adversarial Attack)	2n
Decoder	Conv2D	$16 \times 2n$
	Conv2D+Flattening	$8 \times 2n$
	Dense + ReLU	2M
	Dense + Softmax	2M

4.1 Crafting a White-Box Adversarial Attack with Momentum

Under the framework of the end-to-end communication system, we adopt the white-box and black-box attack crafting method in [8] to craft momentum-based white-box and black-box attacks. Then we compare the performance of the proposed momentum-based attacks with pure jamming attacks and adversarial attacks proposed in [8]. In a white-box attack, we assume the attacker has full knowledge of the autoencoder, including the network structure and parameters; while in a black-box attack, the attacker has no knowledge or limited knowledge of the autoencoder. First of all, Fig. 4 and Fig. 5 show the BLER performance of the MLP autoencoder and CNN autoencoder (described in Table 1 and Table 2) under the proposed momentum-based attacks respectively. We adopt perturbation-to-signal ratio (PSR) as metric, which is the power ratio of the received perturbation to the received signal [9]. To evaluate the effectiveness of the momentum-based adversarial attacks, we also use jamming attacks (Gaussian noise) as counterparts with the same PSR. The values of PSR we selected are –2 dB, –6 dB and –10 dB. Figure 4 and Fig. 5 illustrate that the proposed momentum-based attacks are more destructive than conventional jamming

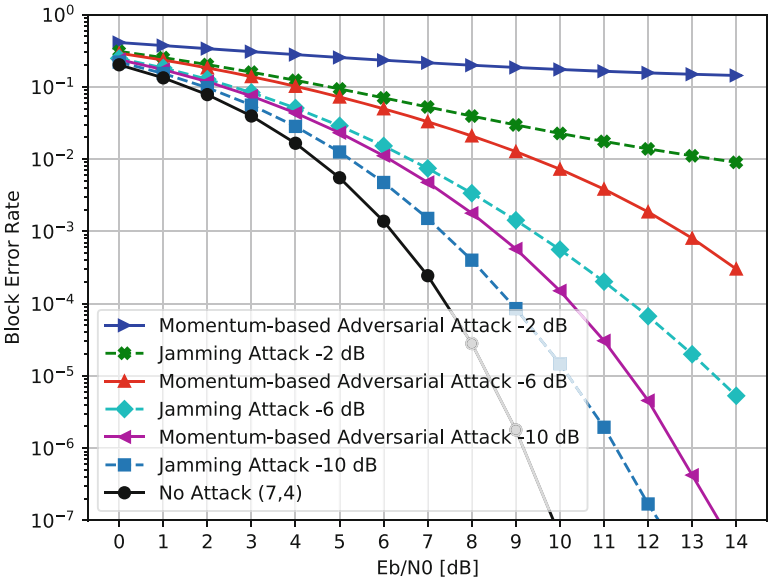

Fig. 4. BLER versus E_b/N_0 under the momentum-based adversarial and jamming attacks against the MLP autoencoder.

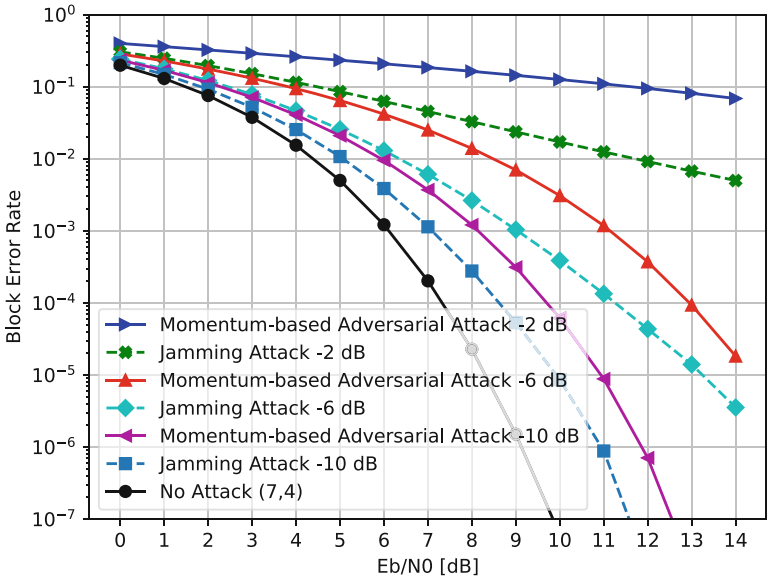

Fig. 5. BLER versus E_b/N_0 under the momentum-based adversarial and jamming attacks against the CNN autoencoder.

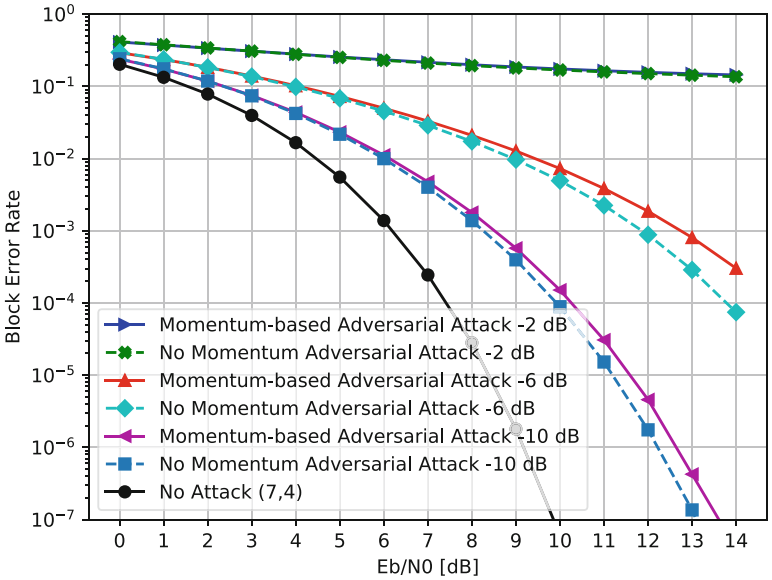

Fig. 6. BLER versus E_b/N_0 under the momentum-based and no momentum white-box attacks.

attacks even for small PSR values, and the performance advantages are become more obvious as the signal-to-ratio (SNR) increases.

Secondly, we compare the performance of the proposed momentum-based method with the adversarial attack method without momentum in Fig. 6. It can be seen that the performance of the proposed momentum-based algorithm is significantly better than the benchmark algorithm in [8] when crafting white-box adversarial attacks. This is due to the momentum-based algorithm accumulates the gradient of the loss function at each iteration to achieve stabilized optimization and get rid of poor local minima and saddle points.

4.2 Crafting a Black-Box Adversarial Attack with Momentum

For verifying the transferability, we also craft black-box attacks for the proposed momentum-based adversarial attack method and the benchmark in [8]. In the black-box attacks, we first use the MLP autoencoder (Table 1) as the substitute model to obtain shift-invariant perturbations as in [8], whereas, based on momentum. Due to the transferability of adversarial attacks, we then attack the CNN autoencoder (Table 2), whose structure and parameters are unknown for the attackers. We use the black-box attack crafting method (without momentum) in [8] as a comparison, and the results of BLER performance are presented in Fig. 7. Comprehensively considering the effect of the white-box and black-box attacks, even the performance improvement of the momentum-based attacks in the black-box attacks are not obvious enough, it can be concluded that the pro-

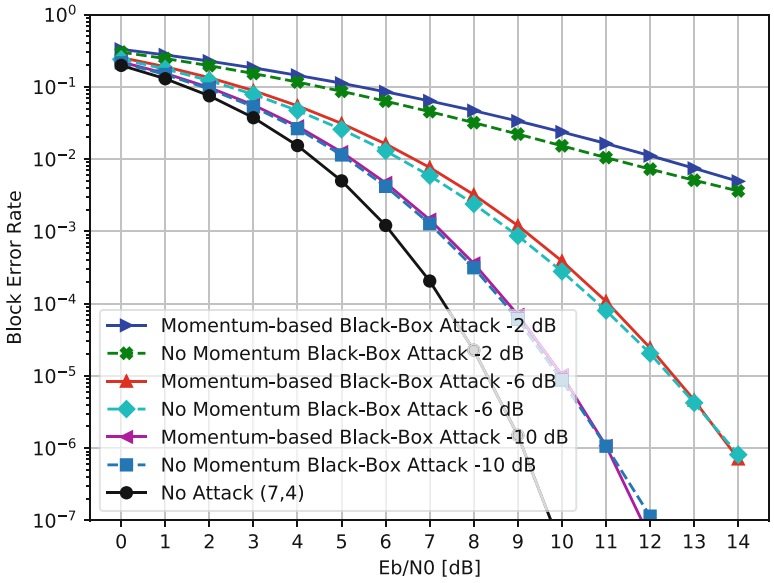

Fig. 7. BLER versus E_b/N_0 under the momentum-based and no momentum black-box attacks.

posed momentum-based method could increase the effectiveness of white-box attacks and alleviate the cost of transferability at the same time.

5 Conclusion

In this paper, we proposed a momentum-based adversarial attack method against the end-to-end communication systems, which incorporates momentum into the adversarial perturbation generative algorithm for better performance. Numerical results showed that, compared with conventional jamming attacks, the adversarial attacks based on momentum can effectively increase the BLER for the DL-based communication models with various network structures. Meanwhile, we also illustrated that the proposed method is more destructive than adversarial attacks without momentum accumulation. Therefore, we demonstrated that the adversarial attack incorporated with momentum can alleviate the trade-off between effectiveness and transferability of the adversarial perturbation, which raises new security issues and robustness evaluation methods for building more reliable end-to-end communication systems.

References

1. O'Shea, T., Hoydis, J.: An introduction to deep learning for the physical layer. IEEE Trans. Cognit. Commun. Netw. **3**(4), 563–575 (2017). https://doi.org/10.1109/TCCN.2017.2758370

2. Dörner, S., Cammerer, S., Hoydis, J., Brink, S.t.: Deep learning based communication over the air. IEEE J. Sel. Top. Signal Process. **12**(1), 132–143 (2018). https://doi.org/10.1109/JSTSP.2017.2784180
3. Ye, H., Liang, L., Li, G.Y., Juang, B.H.: Deep learning-based end-to-end wireless communication systems with conditional gans as unknown channels. IEEE Trans. Wirel. Commun. **19**(5), 3133–3143 (2020). https://doi.org/10.1109/TWC.2020.2970707
4. Chen, X., Cheng, J., Zhang, Z., Wu, L., Dang, J., Wang, J.: Data-rate driven transmission strategies for deep learning-based communication systems. IEEE Trans. Commun. **68**(4), 2129–2142 (2020). https://doi.org/10.1109/TCOMM.2020.2968314
5. Ye, H., Li, G.Y., Juang, B.H.F.: Deep learning based end-to-end wireless communication systems without pilots. IEEE Trans. Cognit. Commun. Netw. 1 (2021). https://doi.org/10.1109/TCCN.2021.3061464
6. Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
7. Szegedy, C., et al.: Intriguing properties of neural networks. arXiv preprint arXiv:1312.6199 (2013)
8. Sadeghi, M., Larsson, E.G.: Physical adversarial attacks against end-to-end autoencoder communication systems. IEEE Commun. Lett. **23**(5), 847–850 (2019). https://doi.org/10.1109/LCOMM.2019.2901469
9. Sadeghi, M., Larsson, E.G.: Adversarial attacks on deep-learning based radio signal classification. IEEE Wirel. Commun. Lett. **8**(1), 213–216 (2019). https://doi.org/10.1109/LWC.2018.2867459
10. Kurakin, A., Goodfellow, I., Bengio, S.: Adversarial machine learning at scale. arXiv preprint arXiv:1611.01236 (2016)
11. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep Learning, vol. 1. MIT press, Cambridge (2016)
12. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. **4**(5), 1–17 (1964)
13. Dong, Y., et al.: Boosting adversarial attacks with momentum. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9185–9193 (2018)
14. Duch, W., Korczak, J.: Optimization and global minimization methods suitable for neural networks. Neural Comput. Surv. **2**, 163–212 (1998)
15. Sutskever, I., Martens, J., Dahl, G., Hinton, G.: On the importance of initialization and momentum in deep learning. In: International Conference on Machine Learning, pp. 1139–1147. PMLR (2013)