Stability-Based Analysis and Defense against Backdoor Attacks on Edge Computing Services

Yi Zhao, Ke Xu, Haiyang Wang, Bo Li, and Ruoxi Jia

Abstract

With the explosive development of mobile Internet and deep learning (DL), intelligent edge computing services based on collaborative learning are widely deployed in various application scenarios. These intelligent services include intelligent applications based on edge computing and DL-based optimization for edge computing (e.g., caching and communicating). However, in a wide variety of domains, DL has been found to be vulnerable to adversarial attacks, especially architecture-independent backdoor attacks. It embeds the attack pattern into the learned model and only performs the attack when it encounters the corresponding trigger. In this article, for the first time we analyze the impact of backdoor attacks on intelligent edge computing services. The simulation results demonstrate that once one or more edge nodes implement backdoor attacks, the embedded attack pattern will rapidly expand to all relevant edge nodes, which poses huge challenges to security-sensitive intelligent edge computing services. Subsequently, we analyze the trade-off between expected performance and ability to defend against backdoor attacks, which sheds new light on designing defense mechanisms for intelligent edge computing services. To address the challenges posed by backdoor attacks, we propose a stability-based defense mechanism. The experimental results demonstrate that the newly proposed defense mechanism can effectively defend against different levels of backdoor attacks without knowing whether there are adversaries, which is conducive to the deployment of the stability-based defense mechanism in real-world scenarios.

Introduction

With the rapid development of mobile and wireless communication technologies, the number of mobile terminals (e.g., iPhone and iPad) and Internet of Things (IoT) devices (e.g., camera and wearable devices) have also grown exponentially. Meanwhile, the performance requirements of these devices are constantly increasing, bringing new pressure on the backbone network.

To alleviate the pressure on the backbone network and support the intelligent service requirements (e.g., storage capacity required for large-scale data and computation resource required for parameter optimization) of mobile terminals and IoT devices, edge computing has become one of the most promising approaches, attracting widespread attention from industry and academia [1, 2].

It has been found that deep learning (DL) methods such as deep neural network (DNN) and deep reinforcement learning (DRL) significantly outperform previous shallow machine learning techniques [3, 4]. In terms of intelligent edge computing services, therefore, a DL-based framework is an indispensable cornerstone. Note that the intelligent edge computing services referred to in this article include not only intelligent applications (e.g., security surveillance [5]) based on edge computing, but also DL-based configuration optimization (e.g., DRL-based edge caching [6]) for edge computing. However, DL models have been found to be vulnerable to adversarial attacks [7, 8]. For example, via adding small pixel-level perturbations to legitimate inputs, an adversary can fool numerous excellent DL models in the field of computer vision. Among multiple attack approaches, architecture-independent backdoor attacks [7] have been considered one of the most promising directions. In terms of implementing this attack, the backdoored DL network has the same architecture as the benign network. Moreover, the attack is only activated when it encounters the corresponding trigger. For benign samples without triggers, the backdoored model works almost normally, thereby avoiding detection. Overall, architecture-independent backdoor attacks are more effective in implementing real-world attacks and bring serious security issues.

Since existing studies mainly focus on centralized learning, one common scenario for implementing this attack is assuming that backdoor samples (i.e., benign samples with embedded triggers) applied to the attack are distributed in the Internet and sneak into training data when the web crawler collects data. However, the probability of having sufficient backdoor samples to sneak into the training data to achieve effective backdoor attacks is low. In fact, intelligent edge computing services are more suitable scenarios for achieving backdoor attacks. One

Yi Zhao is with Tsinghua University, and also with the Beijing National Research Center for Information Science and Technology (BNRist); Ke Xu (corresponding author) is with Tsinghua University, the Beijing National Research Center for Information Science and Technology (BNRist), and Peng Cheng Laboratory (PCL); Haiyang Wang is with the University of Minnesota at Duluth; Bo Li is with the University of Illinois at Urbana-Champaign; Ruoxi Jia is with Virginia Tech.

Digital Object Identifier: 10.1109/MNET.011.2000265

0890-8044/21/$25.00 © 2021 IEEE

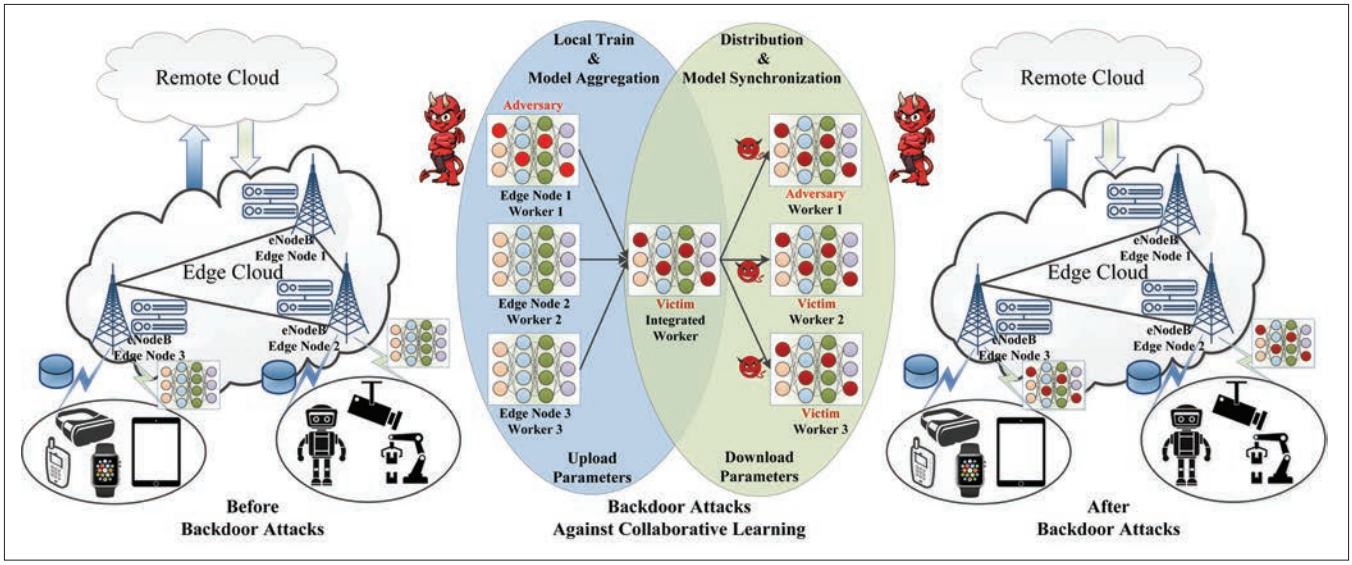

FIGURE 1. Example of backdoor attacks on intelligent edge computing services, where one edge node is an adversary (i.e., training a local model with samples embedded in a backdoor trigger), and the embedded attack pattern is propagated to all edge nodes after the model is aggregated and synchronized.

characteristic of edge computing is that although each edge node can be associated with multiple users, the data that can be obtained is still limited. Collaborative learning can protect the privacy of users associated with a certain edge node while exploiting the value of other nodes' well-trained model. Regarding edge computing, it is common for multiple nodes to share data or models learned from data to achieve collaborative learning [6, 9]. Motivated by this, the intelligent edge computing services we focus on are based on collaborative learning. In such a scenario, once a malicious node (i.e., adversary) exists, or normal nodes are hacked, the attacked model/data can infect other edge nodes, facilitating the propagation of backdoor attacks. Note that in this article, we focus on collaborative edge computing based on model sharing. This is more conducive to the privacy protection of each individual node. As illustrated in Fig. 1, via DL-based collaborative learning, edge computing enables intelligent services for resource-constrained devices (e.g., energy-constrained smartphones, computation-constrained cameras, and storage-constrained watches). Once a malicious edge node (i.e., adversary) exists on the edge cloud, it embeds the attack pattern into the learned DL model. After aggregation and synchronization, other edge nodes will also be attacked to a certain extent. Therefore, the security issues posed by backdoor attacks on edge computing services are more prominent, deserving further analysis.

To the best of our knowledge, we for the first time analyze the impact of backdoor attacks on intelligent edge computing services, and experimentally demonstrate the propagation effect of backdoor attacks among different edge nodes. We propose a stability-based defense mechanism, which sheds new light on robustness enhancement of intelligent edge computing services, achieving effective defense performance against backdoor attacks on edge computing services. More specifically, as formerly noted, collaborative learning in edge computing scenarios is more vulnerable to backdoor attacks in reality. In this article, we utilize the federated learning (FL) framework to achieve collaborative learning for intelligent edge computing services. Then we analyze the propagation of the backdoored model among multiple collaborative workers,[1] as well as the negative effects on intelligent edge computing services. In addition, there are multiple robustness enhancement approaches for a DL model via stability-inducing operations [10], including dropout, regularization, gradient clipping, and so on. From the perspective of stability, we further analyze the trade-off between expected performance and ability to defend against backdoor attacks on DL-based intelligent services. Subsequently, via the theoretical analysis of information theory and simulation experiments, we for the first time propose a stability-based defense mechanism. It can configure the stability-inducing operation parameters of each worker involved in intelligent edge computing services in order to achieve effective defense against backdoor attacks.

The key contributions we make in this article are summarized as follows:

- We for the first time analyze the backdoor attacks on intelligent edge computing services from the perspective of stability, and demonstrate that stability-inducing operations are indispensable modules for defense against backdoor attacks on edge computing services.
- We propose an innovative stability-based defense mechanism for real-world scenarios, which can enable collaborative learning for intelligent edge computing services to effectively defend against backdoor attacks without knowing whether there are adversaries.
- Extensive experimental results demonstrate the potential negative effects of backdoor attacks, and that the proposed defense mechanism can achieve efficient defense performance for intelligent edge computing services.

[1] In terms of edge computing, each participant in collaborative learning refers to one edge node. As far as collaborative learning is concerned, each participant refers more precisely to one worker. In this article, the participant, therefore, is interchangeably represented by edge node or worker.

Related Work

Due to the ubiquity of smart mobile terminals and IoT devices [11], intelligent edge computing services have attracted widespread attention from academia and industry. More specifically, intelligent edge computing services can be divided into two directions: intelligent application service based on edge computing [12, 13] and DL-based configuration optimization service for edge computing [1, 6]. For example, to enable data-driven applications (e.g., image classification) to run on resource-constrained mobile terminals and IoT devices, Wang *et al.* [12] propose an adaptive FL method for resource-constrained edge computing systems. Via the short-delay and high-performance computing services at the edge of the network, Chen *et al.* [14] innovatively propose label-less learning for emotion cognition and deploy it on the edge cloud. With regard to the optimal configuration of edge computing, Wang *et al.* [6] propose a new intelligent optimization framework with mobile edge computing by integrating DRL with FL, which can intelligently optimize mobile edge computing, caching, and communication issues.

As far as intelligent edge computing services are concerned, most of the existing intelligent modules are based on deep neural networks, including DL, DRL, and FL. Despite many surprising advances in deep neural networks, DL is vulnerable to adversarial examples [7, 8] due to the lack of interpretability. For example, Shafahi *et al.* [15] proposes the one-shot kill poisoning attack on transfer learned networks, which can achieve 100 percent attack success rate through injecting one poison instance into the training dataset. One way to achieve backdoor attacks on DL is providing a parallel network to the benign network, which can be applied to detect the backdoor trigger. However, this attack modifies the benign network architecture and is difficult to apply in reality. Gu *et al.* [7] propose the architecture-independent backdoor attack. It embeds the attack pattern into the benign network, which is consistent with the original architecture. As a result, this attack cannot be detected and poses a more serious challenge for DL.

In this article, we for the first time analyze architecture-independent backdoor attacks on intelligent edge computing services. Different from existing studies on backdoor attacks, edge computing focuses on implementing intelligent services through collaborative learning (e.g., FL), which makes it possible for backdoor attacks to spread from some nodes to other nodes. It poses severe challenges for intelligent edge computing services. Moreover, traditional cloud-based collaborative learning has a large number of participants, which can more effectively defend against backdoor attacks from some nodes during model aggregation. In contrast, the number of collaborative learning participants in the edge computing scenario is relatively small; thus, the propagation effect of backdoor attacks will be more obvious. Since there are multiple stability-inducing operations for DL models [10], we for the first time propose a stability-based defense mechanism, which can improve the ability of intelligent edge computing services to defend against backdoor attacks in real-world scenarios.

Backdoor Attacks on Intelligent Edge Computing Services

To fully and concisely analyze the backdoor attacks on intelligent edge computing services, we utilize FL [6, 12] to achieve collaborative learning under the edge computing scenario. More specifically, intelligent edge computing services mainly involve users and edge nodes. As illustrated in Fig. 1, edge nodes refer to base stations with computing and storage capabilities, denoted by $\mathcal{B} = \{B_1, B_2, ..., B_N\}$. Meanwhile, users refer to mobile terminals and IoT devices, denoted by $\mathcal{U} = \{U_1, U_2, ..., U_N\}$. Note that each individual edge node can be associated with multiple users, as illustrated in Fig. 1. As formerly noted, intelligent edge computing services include intelligent application service based on edge computing and DL-based configuration optimization service for edge computing. In the former scenario, resource-constrained users can upload their data to the nearest edge nodes, and then these edge nodes obtain effective DL models through local training and global sharing of model parameters. Finally, the learned models are distributed to users, enabling them to enjoy intelligent services. For each individual edge node, all its associated users are also running the same network architecture, but with different data. Therefore, all users with the same task can be regarded as one user. In the latter scenario, via collaborative learning, edge nodes optimize mobile edge computing, caching, and communication issues. For the DL-based configuration optimization of each individual edge node, the only user is itself. That is, the number of users is the same as the number of edge nodes. Overall, the learning task and the learned model are deployed on each individual edge node. To this end, we use $\mathcal{M} = \{M_1, M_2, ..., M_N\}$ to represent the learned models from these edge nodes.

In terms of FL for edge computing scenarios, the architecture of the model trained by each worker (i.e., edge node) is the same. The difference among workers is local data, which results in different local parameters. After completing local training, the parameters of these different workers will be aggregated together. In this article, the aggregation method we adopt is the average parameter. In addition, the training process is divided into multiple time slots. During time slot t, the parameters of $M_i \in \mathcal{M}$ are denoted by θ_i^t. The aggregated parameters are denoted by $\bar{\theta}^t$. Note that not all workers can update parameters in every time slot, so we use γ to represent the proportion of active workers.

To implement backdoor attacks on intelligent edge computing services, we have to inject backdoor samples (i.e., the legitimate inputs embedded with specific triggers) into the training data. In this article, we focus on targeted backdoor attack. That is, once the adversary successfully infects a victim, the victim will perform a targeted erroneous behavior when encountering an input with a specific trigger. For example, in the case of digital recognition tasks with MNIST, the victim misclassifies each grayscale image with the specific trigger into the correct value plus one and then take the remainder of 10 (i.e., regarding the real label i, the output is $(i + 1)\%$ 10). Obviously, the targeted attacks are more complicated than non-targeted

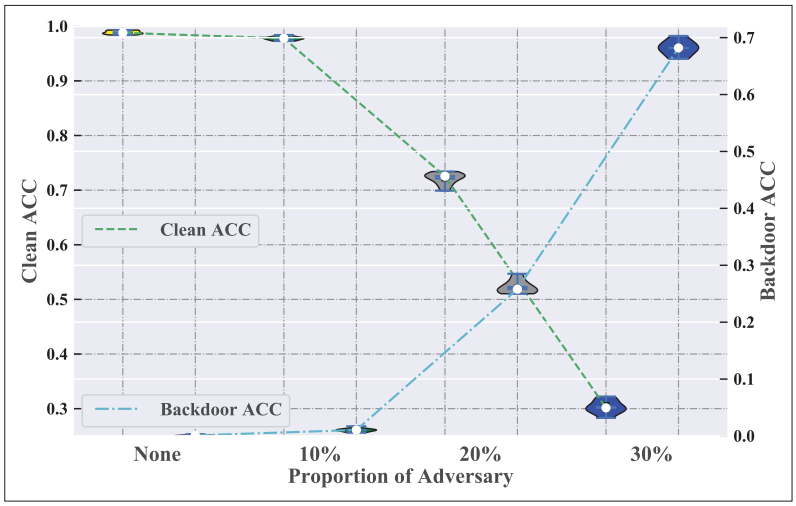

FIGURE 2. Propagation effects of backdoor attacks with different attack proportions in the intelligent edge computing scenario.

Operation	Description
Dropout	Setting a fraction of the gradient weights to zero
Weight decay	Preventing excessive weight growth through penalties
Clipnorm	Clipping gradient norm
Clipvalue	Clipping gradient at specified value
Averaging	Averaging weights of different models

TABLE 1. Major stability-inducing operations.

attacks, posing more challenges to DL models. In the edge computing scenario, backdoor attacks are initiated by some nodes and then spread to other nodes, putting intelligent services at risk. Different from collaborative learning based on cloud computing, the number of participants in collaborative learning under edge computing scenarios is relatively small, and the impact of attacks is more obvious. To demonstrate this, the relevant simulation experiment results can be found in Fig. 2.

As illustrated in Fig. 2, we use distributed digital recognition learning tasks to simulate the propagation effect of backdoor attacks among different edge nodes. More specifically, the simulation experiment includes $N = 10$ edge nodes. The proportion of active workers is $\gamma = 0.8$. *Clean ACC* refers to the probability that the learned model correctly classifies the legitimate inputs, which can be used to measure the expected performance of the model. In contrast, *backdoor ACC* refers to the probability that the learned model misclassifies the legitimate samples with the specific trigger as the corresponding targeted labels, instead of the corresponding real labels. Thus, the *backdoor ACC* can be applied to evaluate the performance of backdoor attacks. A higher value of *backdoor ACC* means a higher probability of successful attack. Both *clean ACC* and *backdoor ACC* are in terms of testing results. The violin plots in Fig. 2 briefly describe the distribution of evaluation indicators (i.e., *clean ACC* and *backdoor ACC*) for 10 workers. For example, we perform experiments with four different attack proportions. It can be found that when there is no adversary, the expected performance of the learned model is good enough, where the *clean ACC* of all workers is higher than 98.4 percent. In addition, since the attack pattern is not embedded, the success rate of the attack is almost zero. However, when there is only one adversary (i.e., 10 percent), the models learned by all workers are attacked. The specific phenomenon declines in *clean ACC*, and *backdoor ACC* is higher than 0. As the proportion of adversaries increases, the expected performance of the learned model gradually decreases, and the probability of successful attacks significantly increases. Moreover, the more discrete distribution of violin appearance indicates that the differences between workers have become more obvious. These results indicate that for edge computing scenarios, once the backdoor attack occurs, it will spread to all participants. Moreover, it will rapidly intensify as the proportion of adversaries increases. Especially in some security-sensitive applications, it will bring huge challenges.

In the experiments involved in Fig. 2, the adopted optimizer is stochastic gradient descent (SGD). For DL models, the optimizer is one of the indispensable modules. The target of the optimizer is to minimize the empirical risk of DL models. Due to its scalability and stability, SGD has been applied in many different domains. It updates the model parameters by repeatedly computing the gradient of the loss function. To make the SGD-based model have sufficient stability (i.e., improving the generalization), there are many stability-inducing operations [10], and some of the major operations are listed in Table 1.

To further clarify the impact of backdoor attacks, we analyze the trade-off between expected performance and ability to defend against backdoor attacks on DL-based intelligent services. Specifically, we use dropout operation as an example for analysis. As described in Table 1, the dropout operation improves the generalization of a DL model through randomly setting a fraction of the gradient weights to zero. In our simulation experiments on the MNIST dataset, we have trained 100 models with different parameters of dropout, and the results can be found in Fig. 3. We can observe the relationship between *backdoor ACC* and *clean ACC* through Fig. 3a. It can be found that as the expected performance (i.e., *clean ACC*) of the model improves, the probability of successful attacks (i.e., *backdoor ACC*) also increases. It is worth noting that there are obvious transition points in Fig. 3a. To the right of the transition point, *backdoor ACC* increases rapidly while *clean ACC* remains almost unchanged. To the left of the transition point, the value of *backdoor ACC* is always at a relatively low level. In addition, Figs. 3b and 3c show the changes in *clean ACC* and *backdoor ACC*, respectively. As the fraction of dropout increases, *clean ACC* gradually decreases. When the fraction of dropout exceeds a certain threshold, *clean ACC* will decrease sharply. A similar trend applies to *backdoor ACC*. The difference is that the thresholds of the transitions corresponding to the two indicators are different. This shows that the impact of stability-inducing operations on different indicators (i.e., *clean ACC* and *backdoor ACC*) is different, which provides an opportunity for the design of the defense mechanism. Through the analysis of Figs. 3a, 3b, and 3c, we can find that an appropriate parameter of dropout can

effectively defend against backdoor attacks. To be precise, this parameter corresponds to the transition point in Fig. 3a. It needs to be restricted to the left of the transition point in Fig. 3b and to the right of the transition point in Fig. 3c.

Stability-Based Defense against Backdoor Attacks on Intelligent Edge Computing Services

Through the previous analysis, we can find that the stability-inducing operations provide an opportunity to defend against backdoor attacks. In this section, we focus on the stability of a DL model and propose the stability-based defense mechanism, enabling intelligent edge computing services to be more suitable for real-world scenarios.

As discussed earlier, the target of intelligent edge computing service is to use some optimization algorithms (back-propagation, SGD, etc.) to find the optimal parameters (i.e., the weights of the neural network). According to information theory, by increasing the uncertainty of parameters during the training phase, the generalization on the testing data can be improved. Stability-inducing operations (e.g., dropout, clipping, and averaging) can improve the generalization of a DL model by discarding part of the optimization information. Based on the analysis of stability by Hardt *et al.* [10], we further define the upper bound of generalization error ε for evaluating the learned model stability. More specifically, based on the *initial benign dataset*, we first construct the *backdoor dataset* with backdoor samples by adding the specific backdoor trigger to some of the samples. In other words, there are both benign and backdoor samples in the *backdoor dataset*. Then we duplicate a completely identical *backdoor dataset* and remove the samples with triggers to construct a new *benign dataset*. We separately train two models, *backdoor model* and *benign model*, with exactly the same network architecture on the referred two datasets (i.e., *backdoor dataset* and *benign dataset*). For each item with index *i* in the *backdoor dataset*, we calculate the loss values through *backdoor model* and *benign model*, respectively. The absolute value of the error between these two loss values is defined as the generalization error of the model (i.e., ε_i). Finally, we use the largest ε_i to represent the upper bound of generalization error (i.e., $\varepsilon = \max\varepsilon_i$).

After defining the upper bound of the generalization error, we take dropout as an example to further analyze the relationship between different stability-inducing parameters and the upper bound of generalization. Note that when defining the stability-based defense mechanism, the definition is not restricted to the operation of dropout. In practice, we can specifically implement our stability-based defense mechanism according to the stability-inducing operations selected by the model. As notified above, ε can be used to reflect the stability of the learned model. A larger ε means that the model stability is relatively low; otherwise, vice versa. Figure 4 shows how the upper bound of generalization error ε changes with different stability-inducing parameters. It can be found that for different models trained with a specific proportion of backdoor samples, the upper bound of generalization error gradually decreases as the parameters increase. In other words, the model is more and more stable, and the ability to defend against backdoor attacks brought by adversarial samples is gradually enhanced, which is consistent with Fig. 3c. Moreover, the upper bound of generalization error also shows a transition point when it changes. Once the parameter exceeds

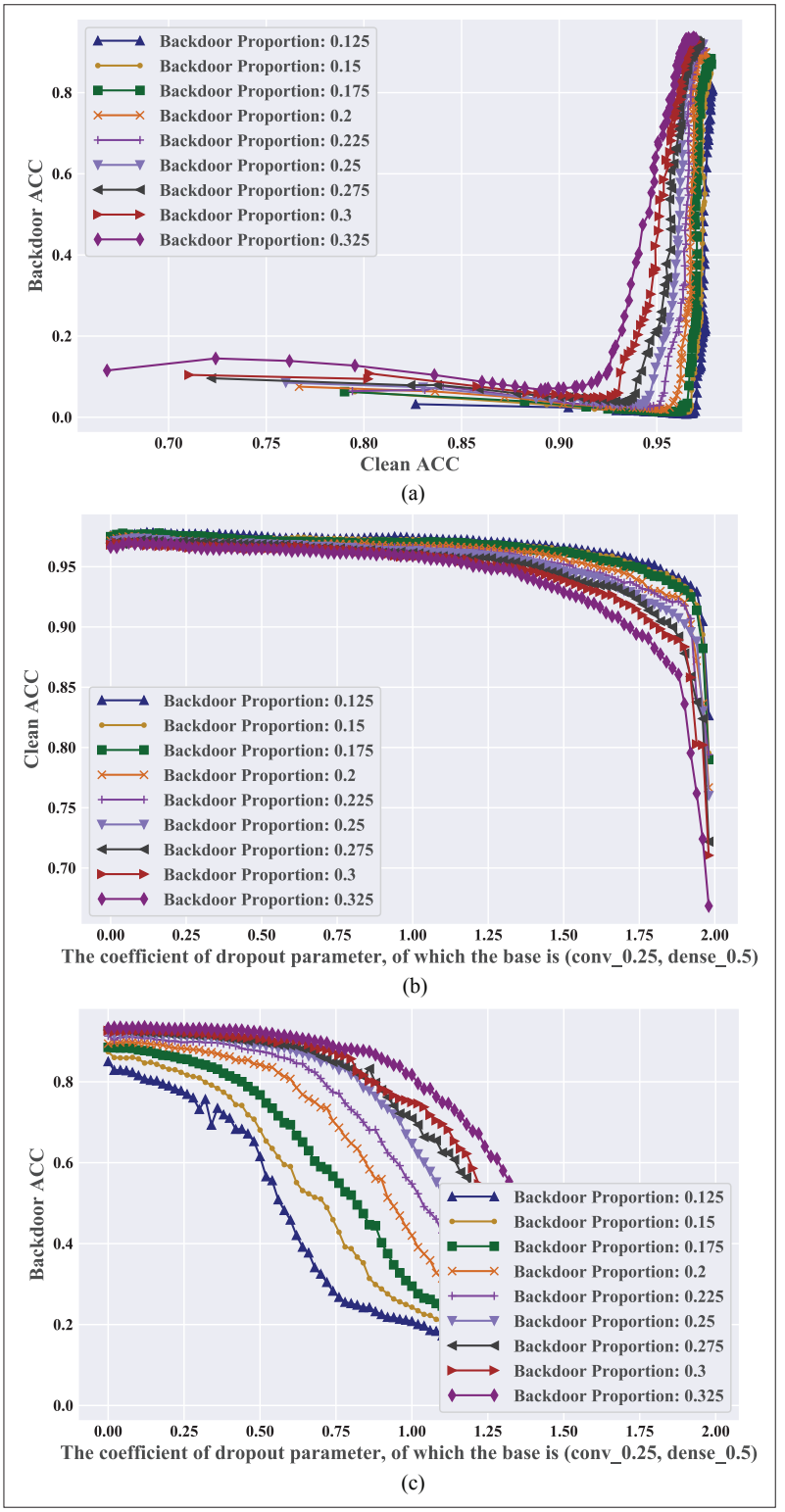

FIGURE 3. Trade-off between collaborative learning performance and ability to defend against backdoor attacks: a) *backdoor ACC* vs. *clean ACC*; b) *clean ACC* with different parameters; c) *backdoor ACC* with different parameters.

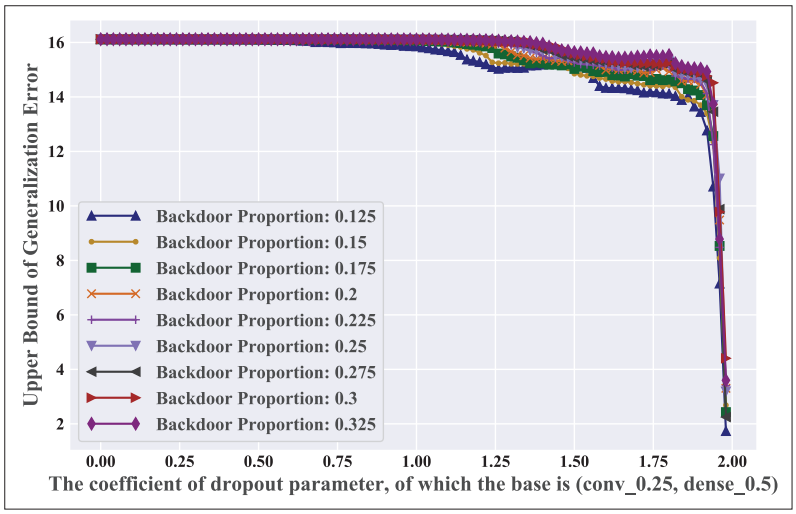

FIGURE 4. Upper bound of generalization error with different parameters.

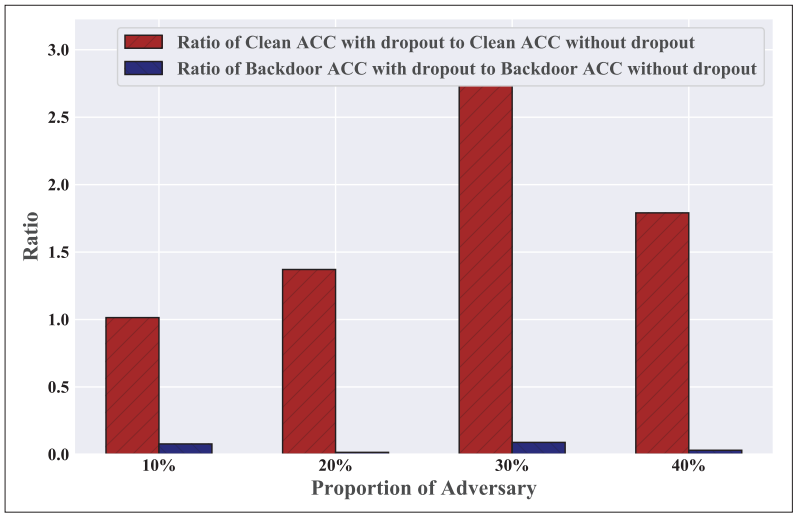

FIGURE 5. Impact of the proposed defense mechanism on different indicators.

the value corresponding to this transition point, the stability of the learned model (i.e., ε) changes dramatically. In addition, for different attack proportions, the upper bound of generalization error changes are very similar (e.g., the decreasing trend and the transition point position).

Through analysis of Figs. 3 and 4, we are clear about the existence of transition points. Note that the transition point will be proved theoretically in future work. Since the change in the upper bound of generalization error, which can be used to reflect the stability of the learned model, is hardly affected by the proportion of backdoor samples, we can use the upper bound of generalization error to find the relevant transition point and then find the optimal stability-inducing parameter for DL models. Regardless of whether or not the dataset is injected with backdoor samples, the parameter corresponding to the transition point of ε is approximate. This meets our requirements for defending against backdoor attacks without knowing whether there are adversaries.

Experiments and Performance Evaluation

To evaluate the performance of the newly proposed stability-based defense mechanism, we conducted simulation experiments on FL-based intelligent edge computing services. More specifically, the reference intelligent edge computing service takes the distributed digital recognition task as an example, which is an important application for intelligent traffic management in real-world scenarios.

Compared to cloud computing, the number of participants in edge computing is relatively small, so our simulation experiments assume $N = 10$ different workers to learn this task collaboratively. For each worker, the training and testing data comes from the MNIST dataset. The data of different workers are independent and identically distributed, and there are no duplicate samples. For those attacked data, the trigger embedded in the legitimate input is a pattern of four pixels, which is consistent with Gu *et al.* [7]. As found in Fig. 4, the transition point for a particular architecture is similar regardless of the proportion of backdoor samples. Therefore, any worker can find the optimal stability-inducing parameter corresponding to the transition point via the upper bound of generalization error. In our simulation, we set the basic parameter of the convolutional layer to 0.25 (abbreviated as conv_0.25), and that of the dense layer to 0.5 (abbreviated as dense_0.5). The coefficients are from [0, 2) with the interval 0.02. The real stability-inducing parameter is the product of coefficient (i.e., the value of X-axis in Fig. 4) and basic parameter. The transition point refers to the fact that the change of the upper bound of generalization error for the first time is greater than four times the average of the previous 15 changes in the upper bounds of generalization error.

Figure 5 shows the evaluation results, where the baseline model is trained without dropout. Note that for *clean ACC* and *backdoor ACC*, the data utilized in Fig. 5 is an average of all 10 worker performance indicators. It can be found that regardless of the proportion of adversaries among all workers, the *backdoor ACC* is almost 0 when we leverage the proposed stability-based defense mechanism to configure stability-inducing parameters. This demonstrates that our proposed mechanism can effectively prevent potential backdoor attacks from different proportions of adversaries. Moreover, the average expected performance (i.e., *clean ACC*) of all 10 collaborating workers has improved compared to the baseline (i.e., without any stability-inducing operations to defend against backdoor attacks). This is mainly because during training, some workers (i.e., adversaries) provide backdoor samples, and the implicit backdoor patterns embedded in the models by these backdoor samples are offset by dropout operation. Therefore, when we calculate the *clean ACC* through benign samples, the expected performance is improved.

Conclusion and Future Work

In this article, we analyze the impact of backdoor attacks on intelligent edge computing services. Through simulation experiments, it has been demonstrated that once some workers become adversaries, all workers will be infected by the malicious models. Moreover, the probability of successful attacks increases significantly with the proportion of adversaries. From the perspective of stability, we further analyze the trade-off between expected performance and ability to

defend against backdoor attacks on DL-based intelligent services. According to the difference of model stability for different indicators, we propose a stability-based defense mechanism to configure stability-inducing parameters. The experimental results demonstrate that the proposed defense mechanism can effectively defend against different levels of backdoor attacks without knowing whether there are adversaries. Moreover, in the case of backdoor samples implicit in the training data, the newly proposed stability-based defense mechanism can also improve the expected performance.

The proposed stability-based defense mechanism sheds new light on robustness enhancement of intelligent edge computing services. In future work, we will further prove the existence of the transition point theoretically. To the best of our knowledge, this is the first article that analyzes how backdoor attacks affect intelligent edge computing services and proposes an effective defense mechanism. Moreover, how to achieve more stable intelligent edge computing services and how to detect backdoor attacks are open issues to study.

Acknowledgments

This work was in part supported by the National Key R&D Program of China with No. 2018YFB0803405, the China National Funds for Distinguished Young Scientists with No. 61825204, the NSFC Project with No. 61932016, the Beijing Outstanding Young Scientist Program with No. BJJWZYJH01201910003011, the Beijing National Research Center for Information Science and Technology (BNRist) with No. BNR2019RC01011, and the PCL Future Greater-Bay Area Network Facilities for Largescale Experiments and Applications with No. LZC0019.

References

[1] M. Chen and Y. Hao, "Task Offloading for Mobile Edge Computing in Software Defined Ultra-Dense Network," *IEEE JSAC*, vol. 36, no. 3, 2018, pp. 587–97.
[2] Y. Hao *et al.*, "Profit Maximization for Video Caching and Processing in Edge Cloud," *IEEE JSAC*, vol. 37, no. 7, 2019, pp. 1632–41.
[3] Y. Zhao *et al.*, "TDFI: Two-Stage Deep Learning Framework for Friendship Inference via Multi-source Information," *Proc. IEEE INFOCOM*, 2019, pp. 1981–89.
[4] R. Wang *et al.*, "DeepNetQoE: Self-Adaptive QoE Optimization Framework of Deep Networks," *IEEE Network*, 2020.
[5] T. Wang *et al.*, "Generative Neural Networks for Anomaly Detection in Crowded Scenes," *IEEE Trans. Info. Forensics and Security*, vol. 14, no. 5, 2018, pp. 1390–99.
[6] X. Wang *et al.*, "In-Edge AI: Intelligentizing Mobile Edge Computing, Caching and Communication by Federated Learning," *IEEE Network*, vol. 33, no. 5, Sept./Oct. 2019, pp. 156–65.
[7] T. Gu *et al.*, "Badnets: Evaluating Backdooring Attacks on Deep Neural Networks," *IEEE Access*, vol. 7, 2019, pp. 47,230–44.
[8] B. Wang *et al.*, "Neural Cleanse: Identifying and Mitigating Backdoor Attacks in Neural Networks," *Proc. IEEE Symp. Security and Privacy*, 2019, pp. 707–23.
[9] B. Wu *et al.*, "Toward Blockchain- Powered Trusted Collaborative Services for Edge-Centric Networks," *IEEE Network*, vol. 34, no. 2, Mar./Apr. 2020, pp. 12–18.
[10] M. Hardt, B. Recht, and Y. Singer, "Train Faster, Generalize Better: Stability of Stochastic Gradient Descent," *Proc. Int'l. Conf. Machine Learning*, 2016, pp. 1225–34.
[11] Y. Zhao *et al.*, "Understand Love of Variety in Wireless Data Market under Sponsored Data Plans," *IEEE JSAC*, vol. 38, no. 4, 2020, pp. 766–81.
[12] S. Wang *et al.*, "Adaptive Federated Learning in Resource Constrained Edge Computing Systems," *IEEE JSAC*, vol. 37, no. 6, 2019, pp. 1205–21.
[13] M. Chen *et al.*, "Living with I-Fabric: Smart Living Powered by Intelligent Fabric and Deep Analytics," *IEEE Network*, vol. 34, no. 5, Sept./Oct. 2020, pp. 156-63.
[14] M. Chen and Y. Hao, "Label-Less Learning for Emotion Cognition," *IEEE Trans. Neural Networks and Learning Systems*, vol. 31, no. 7, 2020, pp. 2430–40.
[15] A. Shafahi *et al.*, "Poison Frogs! Targeted Clean-Label Poisoning Attacks on Neural Networks," *Proc. Advances in Neural Info. Processing Systems*, 2018, pp. 6103–13.

Biographies

Yi Zhao [S'19] received his B. Eng. degree from the School of Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China, in 2016. Currently, he is pursuing a Ph.D. degree in the Department of Computer Science and Technology at Tsinghua University, Beijing, China. His research interests include network economics, network security, machine learning, social network, and game theory. He is a Student Member of ACM.

Ke Xu [M'02, SM'09] received his Ph.D. from the Department of Computer Science and Technology at Tsinghua University, where he serves as a full professor. He has published more than 200 technical papers and holds 11 U.S. patents in the research areas of next-generation Internet, blockchain systems, the Internet of Things, and network security. He is a member of ACM. He has guest edited several Special Issues in IEEE and Springer journals. He is an Editor of the *IEEE Internet of Things Journal*. He is Steering Committee Chair of IEEE/ACM IWQoS.

Haiyang Wang [S'08, M'13] received his Ph.D. degree in computing science from Simon Fraser University, Burnaby, British Columbia, Canada, in 2013. He is currently an associate professor with the Department of Computer Science, University of Minnesota at Duluth. His research interests include cloud computing, peer-to-peer networking, social networking, big data, and multimedia communications.

Bo Li received her Ph.D. degree from Vanderbilt University in 2016. She is currently an assistant professor with the Computer Science Department, University of Illinois at Urbana-Champaign. Her research focuses on machine learning, security, privacy, and game theory. She is the recipient of an MIT Technology Review TR-35 award and a Symantec Fellowship award.

Ruoxi Jia [S'15, M'20] received her Ph.D. degree from the University of California Berkeley in 2018. She is currently an assistant professor in the the Bradley Department of Electrical and Computer Engineering at Virginia Tech. Her research interests lie broadly in the span of machine learning, security, privacy, and cyber-physical systems.