MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://mmc.committees.comsoc.org/

MMTC Communications – Review

IEEE COMMUNICATIONS SOCIETY

Vol. 14, No. 2, April 2023

TABLE OF CONTENTS

Message from the Review Board Directors

Welcome to the April 2023 issue of the IEEE ComSoc MMTC Communications – Review.

This issue comprises four reviews that cover multiple facets of multimedia communication research including image enhancement, mobile edge computing, machine learning, and vehicular networks. These reviews are briefly introduced below.

The first paper, published in IEEE Journal on Selected Areas in Communications and edited by Dr. Takuya Fujihashi, proposes a deep learning (DL)-enabled semantic communication system for speech signals to solve issues of the traditional communication system facing about performance.

The second paper, edited by Dr. Guitao Cao, was published in IEEE Transactions on Multimedia. This paper investigates an end-to-end multiscale learning network for recognition of emotions in images.

The third paper, edited by Dr. Qichao Xu, was published in IEEE Trans. Vehicular Technology. The authors proposed a UAV-assisted cloud-edge system to cooperate with the cloud center to provide cache and computing services for the ground users. A novel algorithm is presented to jointly optimize the caching and offloading decision schemes.

The fourth paper, is published in IEEE Wireless Communications, and edited by Dr. Shengjie Xu. The paper studied backdoor attacks and defenses in federated learning.

All the authors, reviewers, editors, and others who contribute to the release of this issue deserve appreciation with thanks.

IEEE ComSoc MMTC Communications – Review Directors

Yao Liu
Rutgers University, USA
Email: yao.liu@rutgers.edu

Wenming Cao
Shenzhen University, China
Email: wmcao@szu.edu.cn

Dongfeng (Phoenix) Fang
California Polytechnic State University, USA
Email: dofang@calpoly.edu

Ye Liu
Macau University of Science and Technology, Macau, China Email: liuye@must.edu.mo

Semantic Communication for Speech Information

A short review for "Semantic Communication Systems for Speech Transmission"

Edited by Takuya Fujihashi

Z. Weng and Z. Qin, "Semantic Communication Systems for Speech Transmission," IEEE Journal on Selected Areas in Communications, Vol. 39, No. 8, Aug. 2021.

Deep Learning (DL) is one of promising techniques for communications to improve throughput and system performance. For example, some DL solutions have designed for physical layer [1], resource allocation [2], and other issues to solve the existing technical problems.

Such DL-based communication system merges one or multiple communication modules of the traditional block-wise architecture by using a neural network (NN) to realize the intelligent transceiver. However, the state-of-the-art models mainly focus on performance improvement at the bit or symbol level, which usually takes bit-error rate (BER) or symbol-error rate (SER) as the performance metric. Specifically, the major task in the traditional communication systems and the developed DL-based communication systems is to accurately recover the transmitted message, which is represented by digital bit sequences. Such communications are no longer ideal since the digital bit sequences contain information that could be not to the intelligent tasks at the receiver. In addition, in the typical communication systems, the amount of transmitted data is generally larger than the required data and it limits the number of devices to be covered by the same network bandwidth.

Motivated by the issues of the traditional communications, some studies exploit DL architectures for the realization of semantic communications [3, 4]. In the semantic communication systems, the transmitter extracts the semantic information from the source and only transmits the semantic information, and the receiver recovers the information via minimizing the semantic error instead of BER and SER. Here, the semantic information represents the information relevant to the transmission goal at the receiver. Note that even the most cutting-edge studies do not define the exact meaning of the semantic information by a mathematical formula. For the realization, some studies have designed end-to-end (E2E) semantic communication systems to address the bottlenecks in traditional block-wise communication systems. The existing E2E semantic communication systems mainly focus on the image and text transmission, whereas this paper aims at semantic communication for speech signals. Specifically, the authors proposed a DL-enabled semantic communication system for speech signals, named DeepSC-S.

The contributions of the proposed DeepSC-S are three-fold. The first one is to design semantic transmitter and receiver using individual NNs. The transmitter side consists of individual components: the semantic source encoder and the channel encoder, each component is implemented by an independent NN. Here, the input for the transmitter side is speech sample sequence. The receiver side also consists of individual components of the semantic source decoder and channel decoder. Specifically, an attention-based two-dimension (2D) convolutional neural network (CNN) is used for the semantic encoder/decoder and a 2D CNN is adopted for the channel encoder/decoder. Although the traditional communications usually use the advanced channel coding and the bit-to-symbol mapping for error protection, the proposed DeepSC-S directly maps the output of the channel encoder to transmission symbols. The proposed DeepSC-S trains both semantic transmitter and receiver using the loss function of mean square error (MSE) to reconstruct the raw speech signals via wireless channels.

The second contribution is to employ a squeeze-and-excitation (SE) network [5] in the semantic source encoder and decoder to extract essential information from the source and assign high values to the essential information for setting the importance of the essential information during the training phase. The attention mechanism based on the SE network, named SE-ResNet, is also combined with the residual network to alleviate the problem of gradient vanishing. The proposed SE-ResNet can improve the efficiency of the semantic communications. Here, multiple SE-ResNet modules can be sequentially connected

considering a trade-off between the efficiency and complexity during the training phase.

The third contribution is to evaluate the effectiveness of the proposed DeepSC-S under the different performance metrics and wireless channel environments. Authors employed the signal-to-distortion ration (SDR) [6] and perceptual evaluation of speech distortion (PESQ) [7] as the performance metrics to measure the perceptual degradation of speech signals. The higher SDR represents that the speech information is recovered with better quality, i.e., easier to understand for human beings. PESQ considers the quality of speech signals under various conditions e.g., background noise, analog filtering, and variable delay. In view of the wireless channel environments, authors considered additive white Gaussian noise (AWGN), Rayleigh, and Rician channels.

References:

[1] H. Ye, G. Y. Li, and B. H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," IEEE Wireless Communication Letter, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[2] L. Liang, H. Ye, G. Yu, and G. Y. Li, "Deep-learning-based wireless resource allocation with application to vehicular networks," Proceedings of IEEE, vol. 108, no. 2, pp. 341–356, Feb. 2020.

[3] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, "Deep joint source-channel coding for wireless image transmission," IEEE Transactions on Cognitive Communications and Networking, vol. 5, no. 3, pp. 567–579, Sep. 2019.

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

[5] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[6] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," Transactions on Audio, Speech, and Language Processing, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.

[7] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs," in Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing, 2001, pp. 749–752.

Takuya Fujihashi received the B.E. degree in 2012 and the M.S. degree in 2013 from Shizuoka University, Japan. In 2016, he received Ph.D. degree from the Graduate School of Information Science and Technology, Osaka University, Japan. He is currently an assistant professor at the Graduate School of Information Science and Technology, Osaka University since April, 2019. He was an assistant professor at the Graduate School of Science and Engineering, Ehime University, Japan from Jan. 2017 to Mar. 2019. He was research fellow (PD) of Japan Society for the Promotion of Science in 2016. From 2014 to 2016, he was research fellow (DC1) of Japan Society for the Promotion of Science. From 2014 to 2015, he was an intern at Mitsubishi Electric Research Labs. (MERL) working with the Electronics and Communications group. He selected one of the Best Paper candidates in IEEE ICME (International Conference on Multimedia and Expo) 2012. His research interests are in the area of video compression and communications, with a focus on multi-view video coding and streaming.

Multiscale Emotion Representation Learning for Affective Image Recognition

A short review for "Multiscale Emotion Representation Learning for Affective Image Recognition"

Edited by *Guitao Cao*

H. Zhang and M. Xu, "Multiscale Emotion Representation Learning for Affective Image Recognition," in IEEE Transactions on Multimedia, doi: 10.1109/TMM.2022.3144804.

Recognition of emotions in images is a critical aspect of image understanding [1]. This interdisciplinary field combines knowledge from fields such as visual cognition theory, psychology, and pattern recognition to investigate the emotional responses of humans to images. The applications of this field are wide-ranging and include entertainment [2], opinion mining [3], and affective image retrieval [4], among others.

Compared to other tasks in image understanding, emotion recognition in images presents unique challenges due to high intra-class variations [5]. This is because images conveying the same emotion can be taken in very different scenes with various objects, making it difficult to learn robust emotional representations. Typically, the dominant emotion in an image is localized to a specific region, while other areas of the image may exhibit a neutral or non-dominant emotion. Recent research has shown that incorporating local affective regions can improve recognition performance [6, 7]. However, one major drawback of these studies is the need for emotion region annotations, which can be both time-consuming to obtain and computationally demanding to discover. In addition, the potential benefits of learning multiscale emotion features for affective image recognition tasks have not yet been thoroughly explored by existing studies.

In this paper, the author's main contributions are in two aspects: in response to the limitations of existing datasets for image emotion recognition that these datasets only contain image-level annotations, the authors of this study have proposed an end-to-end multiscale learning network that includes an affective region detection module and a multiscale learning module for recognizing emotions in images. Focusing on the investigation that emotion clues in an image can be found from multiple scales, the authors introduce a kernel-based graph attention network.

To be specific, the proposed network includes two steps. In the first step, called pseudo affective region generation, the network avoids the need for manual annotation of affective regions by employing a weakly supervised approach to extract pseudo affective regions using the CAM method [8]. This step adopts a threshold procedure to obtain the bounding box from a class activation map, which is only performed for class activation maps corresponding to the true labels. The proposed network performs binarization for the class activation map with a threshold value that equals 20% of the map's maximum value and takes the bounding box covering the largest connected component in the binarized map. The resulting bounding box regions represent the most distinctive part of the images that convey an emotion. Once the bounding boxes are obtained, the pseudo regions are used to train the affective region detection module, which is designed to identify local regions that elicit human emotions.

In the second step, the proposed multiscale learning network adopts a two-stage architecture similar to the Faster RCNN framework [9]. In the first stage, the network leverages the affective region detection module to identify local affective regions. The affective region detection module adopts the Faster RCNN framework as backbone network. It uses a region proposal network to generate candidate regions, and then applies a fast RCNN detector to determine whether a candidate region is an affective region. Once an affective region is detected, the second stage of the proposed network extracts three-scale features and encodes them using a kernel-based graph attention network. This network computes attention weights by comparing similarities in the reproducing kernel Hilbert space (RKHS). Finally, the three-scale features are concatenated for emotion classification.

The proposed network was extensively tested on various benchmark datasets for both emotion recognition and binary sentiment classification. The experimental results demonstrate that the network outperformed previous state-of-the-art models, achieving comparable or improved

performance. The results suggest that the proposed network is highly effective in accurately recognizing and classifying emotions.

In summary, the proposed end-to-end multiscale learning network for recognition of emotions in images is a two-stage architecture. The first stage utilizes the CAM method to generate pseudo affective regions, which are then used to train the proposed network for affective region detection. By integrating features from the broad context, the proposed network can learn improved emotion representations. To further enhance the performance, this paper introduces a kernel-based graph attention network that encodes features from different scales. The results of the ablation studies show that the kernel-based attention is effective in improving the recognition performance, surpassing the conventional dot-product attention. The proposed network is evaluated on different benchmark datasets for multiclass emotion recognition and binary sentiment classification. The experimental results demonstrate that the proposed network achieves improved or comparable performance compared to previous state-of-the-art methods.

References:

[1] D. She, J. Yang, M.-M. Cheng, Y.-K. Lai, P. L. Rosin, and L. Wang, "Wscnet: Weakly supervised coupled networks for visual sentiment classification and detection," IEEE Transactions on Multimedia, vol. 22, no. 5, pp. 1358-1371, 2019.
[2] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang, "Can we understand van gogh's mood? learning to infer affects from images in social networks," in Proceedings of the 20th ACM international conference on Multimedia, 2012, pp. 857-860.
[3] Q. You, J. Luo, H. Jin, and J. Yang, "Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia," in Proceedings of the Ninth ACM international conference on Web search and data mining, 2016, pp. 13-22.
[4] L. Pang, S. Zhu, and C.-W. Ngo, "Deep multimodal learning for affective analysis and retrieval," IEEE Transactions on Multimedia, vol. 17, no. 11, pp. 2008-2020, 2015.
[5] H. Zhang and M. Xu, "Recognition of emotions in user-generated videos with kernelized features," IEEE Transactions on Multimedia, vol. 20, no. 10, pp. 2824-2835, 2018.
[6] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," Neurocomputing, vol. 333, pp. 429-439, 2019.
[7] J. Zhang, X. Liu, M. Chen, Q. Ye, and Z. Wang, "Image sentiment classification via multi-level sentiment region correlation analysis," Neurocomputing, vol. 469, pp. 221-233, 2022.
[8] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921-2929.
[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," Advances in neural information processing systems, vol. 28, 2015.

Guitao Cao obtained her Ph.D. in 2006 from Shanghai Jiao Tong University with a focus on pattern recognition and intelligence system. She is currently a professor of Software Engineering Institute, East China Normal University (ECNU), Shanghai, China. ECNU is the top tier university in China with a high rank (Level A) in Software Engineering in China. She was also a visiting researcher with University of Missouri at Columbia. She has published decades of peer reviewed papers in top venues including IEEE Transactions on Cybernetics, IEEE Transactions on Multimedia, and IEEE Transactions on Biomedical Engineering. Prof. Cao is also the Principal Investigator for many research funding with major sponsors including the National Science Foundation of China, Ministry of Industry and Information Technology of the People's Republic of China, and Science Foundation of Shanghai. Her research interests include pattern recognition, image processing and machine learning

Computation Offloading and Caching in UAV-assisted Cloud-edge System

A short review for "Computation Offloading and Resource Allocation in Unmanned Aerial Vehicle Networks"

Edited by Qichao Xu

B. Liu, C. Liu, and M. Peng, "Computation Offloading and Resource Allocation in Unmanned Aerial Vehicle Networks," in IEEE Trans. Vehicular Technology, vol. 72, no. 4, pp. 4981–4995, Apr. 2023.

Unmanned aerial vehicles (UAVs) have recently drawn extensive research attention because of the mobility, low environmental requirements and multiple sizes. Compared with conventional mobile edge computing (MEC) systems, UAV-assisted MEC systems can provide better flexible and on-demand services, where the UAVs act as MEC servers that respond more quickly to user requests through high-quality line-of-sight (LoS) communication links.

However, a significant obstacle to the actual application of UAV-assisted MEC networks is their constrained computation and energy storage capabilities. One possible solution to this problem is to use cloud computing as a complement to edge UAVs (EUAVs). In such multi-tier cloud-edge combination network, the cloud center has powerful computing resources and a large amount of cache capacity, and EUAV can be connected to these networks through fronthaul links, thus making up for the shortcomings of UAV-assisted MEC in achieving cost-effective computing with delay satisfaction.

UAV-assisted computing [1] and caching [2] have been discussed separately in the current work. Limited research work has taken both into account. A caching mechanism introduced into computing can significantly improve resource utilization, reduce latency, and improve user experience by preventing the duplicate transmission of popular file contents and handling the burst traffic promptly. Therefore, it is worthwhile to research how to schedule tasks between the edge and the cloud by utilizing the relationship between caching and computation, as well as how to balance the latency and energy consumption of the UAV-assisted cloud-edge system. The selection and offloading of cached data face challenge due to the uneven user demands and heterogeneity of the cloud-edge architecture. In addition, the above work mainly considers the use of convex optimization methods [1, 2, 3] or deep reinforcement learning (DRL) methods [4, 5].

This paper applies the caching and computing functions of edge servers to consider UAV-assisted cloud-edge networks, in which EUAV as MEC servers cooperates with cloud centers to provide caching and computing services for ground users. UAV-assisted cloud-edge network consists of a cloud layer, an edge layer, and a terminal layer. UAVs with constrained cache and computing capabilities are deployed as MEC servers in the edge layer close to users. In the cloud layer, a tethered UAV (TUAV) physically wired to a ground control center (GCC) of power caching and computing capabilities is deployed.

The objective of this paper is to jointly optimize the caching and offloading decision schemes. As a result, this study formulates the cost minimization problem and reveals its dependence on caching and offloading decisions, resource allocation, as well as UAV deployment strategies. A hybrid framework is proposed for problem-solving that minimizes the weighted sum cost of latency and energy consumption of the networks under consideration while inducing relatively low computational complexity.

Specifically, the optimization problem is decomposed into three sub-problems in this article: the EUAV deployment sub-problem, the resource allocation sub-problem, and the caching and offloading decision sub-problem.

For the EUAV deployment sub-problem, the authors suggested a sequential convex programming (SCP) algorithm with the given caching and offloading decision scheme and the

resource allocation scheme. After converting the objective function from a non-convex function to a convex optimization problem, the suggested SCP-based EUAV deployment algorithm uses the CVX solver to resolve this subproblem.

For the resource allocation sub-problem, the objective function is non-convex and challenging to directly solve with the given caching and offloading decision scheme and the EUAV deployment scheme. Both the objective function and the constraint contain nonlinear functions after being transformed into a convex function, making it a nonlinear programming issue which the SQP algorithm can resolve. By suggesting a resource allocation algorithm based on SQP, this paper resolves the resource allocation subproblem.

For the caching and offloading decision sub-problem, the objective problem is NP-hard. The authors proposed a modified DQN-based algorithm to carry out caching and offloading decisions. An SS-DQN algorithm for caching and offloading choices is proposed. The benefits of the DRL method and the convex optimization method are combined in this hybrid approach. The SS-DQN algorithm only needs to train the caching and offloading decision variables instead of all the variables. This method drastically reduces the corresponding state-action space and effectively enhances the executive efficiency of the DQN algorithm.

Numerical results show significant performance gain compared to the existing edge computing only schemes and illustrated the advantages of combining DRL and convex optimization to improve the cost performance and the training efficiency.

In summary, the authors considered a UAV-assisted cloud-edge system to cooperate with the cloud center to provide cache and computing services for the ground users. A novel SS-DQN algorithm is put up to address the NP-hard problem. It combines the benefits of the SCP, SQP, and DQN methods. The proposed algorithm allowed the system to learn the optimal parameter tuple that minimize the system cost in a more efficient way.

References:

[1] H. Mei, K. Yang, Q. Liu, and K. Wang, "Joint trajectory-resource optimization in UAV-enabled edge-cloud system with virtualized mobile clone," IEEE Internet Things J., vol. 7, no. 7, pp. 5906–5921, Jul. 2020.

[2] A. A. Nasir, "Latency optimization of UAV-enabled MEC system for virtual reality applications under rician fading channels," IEEE Commun. Lett., vol. 10, no. 8, pp. 1633–1637, Aug. 2021.

[3] Z. Yu, Y. Gong, S. Gong, and Y. Guo, "Joint task offloading and resource allocation in UAV-enabled mobile edge computing," IEEE Internet Things J., vol. 7, no. 4, pp. 3147–3159, Apr. 2020.

[4] A. Sacco, F. Esposito, G. Marchetto, and P. Montuschi, "Sustainable task offloading in UAV networks via multi-agent reinforcement learning," IEEE Trans. Veh. Technol., vol. 70, no. 5, pp. 5003–5015, May 2021.

[5] G. Faraci, C. Grasso, and G. Schembra, "Design of a 5G network slice extension with MEC UAVs managed with reinforcement learning," IEEE J. Sel. Areas Commun., vol. 38, no. 10, pp. 2356–2371, Oct. 2020.

Qichao Xu received Ph.D degree from the school of Mechatronic Engineering and Automation, Shanghai University, Shanghai, P. R. China, in 2019. He is currently an associate professor with Shanghai university. His research interests are in trust and security, the general area of wireless network architecture, internet of things, vehicular networks, and resource allocation. He has published more than 50 papers in some respected journals, e.g., IEEE TIFS, IEEE TDSC, IEEE TWC, IEEE TII, IEEE TVT, etc. He was receipt of the best paper awards from several international conferences including IEEE IWCMC2022, IEEE MSN2020, EAI MONAMI2020, IEEE Comsoc GCCTC2018, IEEE CyberSciTech 2017, and WiCon2016.

Backdoor Attacks and Defenses in Federated Learning

A short review for "Backdoor Attacks and Defenses in Federated Learning"

Edited by Shengjie Xu

X. Gong, Y. Chen, Q. Wang, and W. Kong, "Backdoor Attacks and Defenses in Federated Learning: State-of-the-Art, Taxonomy, and Future Directions," in IEEE Wireless Communications, vol. 30, no. 2, pp. 114-121, April 2023, doi: 10.1109/MWC.017.2100714.

Federated learning was first proposed by Google in 2016. Its key idea is to train a sophisticated global model on datasets distributed across various participants while protecting privacy. In federated learning, there are two parties: a central cloud server and various participants. Each participant maintains a local model updated using the local training data. The central server maintains the global model aggregated from submitted local models.

Backdoor attacks aim to mislead the backdoored model to exhibit abnormal behavior on any sample stamped with the backdoor trigger but behave normally on all benign samples. It first appeared in centralized learning and has extended to federated learning in recent years.

Regarding the attack goal, the existing backdoor attacks include untargeted and targeted attacks. Untargeted backdoor attacks only aim to damage the main task accuracy of the global model, while the goal of targeted backdoor attacks is to misclassify all the backdoored samples to the specific target label. Regarding the number of triggers, the existing backdoor attacks consist of single-trigger backdoor attacks and multi-trigger backdoor attacks. Note that different triggers usually target different targeted labels.

Through polluting the training dataset, the attacker can realize the attack goal. Specifically, given the backdoor trigger, the attacker generates an (input, label) pair for every selected training sample. One is the original training data sample and its corresponding ground-truth label, and the other is the backdoored data sample with the trigger and the targeted label. To reduce the impact on model performance, the adversary only selects a small part of training data to construct the poisoned dataset. After training on that poisoned dataset, the model will be backdoored.

Backdoor attacks happen during the training phase since attackers need to manipulate the training process of the DNN by poisoning the training dataset. Moreover, in contrast to typical adversarial examples that customize noises for each sample, backdoor attacks generate a universal backdoor trigger that can be added to any sample and trigger the backdoor.

The training phase consists of two sub-phases: training data collection and learning procedure. Training data collection is to gather a training dataset, and the learning procedure generates a model based on the training dataset. According to the attack stage, we classify the existing backdoor attacks against federated learning into two categories: data poisoning attacks and model poisoning attacks.

Byzantine-robust aggregation (e.g., Krum, mean, trimmed mean) can be used for mitigating Byzantine attacks against federated learning. However, most of these aggregation methodologies assume independent and identically distributed (i.i.d.) data, thus, failing to defend existing backdoor attacks [1, 2] (most of them assume a non-i.i.d. scenarios). Moreover, these strategies do not differentiate backdoored updates from the benign ones. They only want to tolerate the attacks and alleviate the malicious effects.

Recently, facing the severe impacts of backdoor attacks, many tailored defense works have been proposed. As far as we know, the state-of-the-art countermeasures against backdoor attacks in federated learning can be classified into three categories: anomaly update detection, robust

federated training, and backdoored model restoration.

The paper also presents a comprehensive comparison of the existing backdoor attacks and defenses in federated learning. Performance evaluation is conducted through experiments, as both attacks' and defenses' source codes are not available. The evaluation metrics are attack success rate (ASR) and main task accuracy (MTA), where MTA evaluates the prediction accuracy of the backdoored model on benign samples. ASR is calculated as the probability that a backdoored sample is misclassified to the target label.

Four future research on attacks aspects that are worth exploring. First of all, most of the current backdoor attacks are targeted at horizontal federated learning, where datasets have the same feature space yet different from each other. However, vertical federated learning is also widely used in the industry. It is more challenging since the adversaries usually need to learn about labels in such a scenario. How to design more effective backdoor attacks against vertical federated learning is a potential research direction.

Second, it is also necessary to design invisible backdoor attacks against federated learning. In the training phase of federated learning, the attacker can directly inject visible poisoned samples into the local training dataset since both the server and other clients cannot inspect that private dataset. However, after the backdoored global model is deployed on the user devices, the visible trigger will raise suspicion.

Third, similar to centralized learning, studying physically practical backdoor attacks is also imminent. Although in the digital setting, most of the existing backdoor attacks can achieve a high attack success rate, the trigger will be influenced by a variety of factors, such as noise, lighting, and blurring in the physical world.

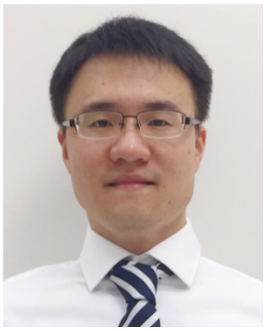
Finally, all the existing backdoor attacks against federated learning use random triggers (e.g., sticker, random pixel perturbation) to inject the backdoors. Four future research on defense aspects is worth exploring. Designing more effective defenses against backdoor attacks while preserving model accuracy and protecting privacy is a future direction. Second, it is critical to design defense strategies for other tasks besides the image domain, for example, voice, text. Third, defense works also need to be improved to adapt to those single-trigger backdoor attacks. Last, only one post-training defense work focuses on repairing the backdoored model [3].

References:

[1] C. Xie et al., "DBA: Distributed Backdoor Attackst Federated Learning," Proc. Int'l. Conf. Learning Representations, 2019.

[2] Y. H. D. E. Eugene Bagdasaryan et al., "How to Backdoor Federated Learning," Proc. Int'l. Conf. Artifi cial Intelligence and Statistic, 2020, pp. 2938–48.

[3] C. Wu et al., "Mitigating Backdoor Attacks in Federated Learning," arXiv preprint arXiv:2011.01767, 2020.



Shengjie Xu [SM'14-M'19] received a Ph.D. degree in Computer Engineering from the University of Nebraska-Lincoln and an M.S. degree in Telecommunications from the University of Pittsburgh. Before that, he held a B.E. degree in Computer Science and Information Security. He is an assistant professor in the Management Information Systems Department at the Fowler College of Business at San Diego State University. His research interests are cybersecurity, trustworthy AI and robust machine learning, secure edge computing, and critical infrastructure protection. He serves as a Technical Editor for IEEE Wireless Communications. He is the recipient of the IET Journals Premium Award for Best Paper. He holds multiple professional certifications in cybersecurity and computer networking.

Multimedia Communications Technical Committee Officers

Chair: Chonggang Wang, InterDigital, USA
Steering Committee Chairs: Shaoen Wu, Illinois State University, USA
Abderrahim Benslimane, University of Avignon, France
Vice Chair – America: Wei Wang, San Diego State University, USA
Vice Chair – Asia: Liang Zhou, Nanjing University of Post and Telecommunications, China
Vice Chair – Europe: Reza Malekian, Malmö University, Sweden
Letters & Member Communications: Qing Yang, University of North Texas, USA
Secretary: Han Hu, Beijing Institute of Technology, China
Standard Liaison: Weiyi Zhang, AT&T Research, USA

MMTC examines systems, applications, services and techniques in which two or more media are used in the same session. These media include, but are not restricted to, voice, video, image, music, data, and executable code. The scope of the committee includes conversational, presentational, and transactional applications and the underlying networking systems to support them.