Federated Learning for 6G Communications: Challenges, Methods, and Future Directions

This is an invited paper.

Yi Liu[1], Xingliang Yuan[2], Zehui Xiong[3], Jiawen Kang[4,*], Xiaofei Wang[5], Dusit Niyato[6]

[1] School of Data Science of Technology, Heilongjiang University, Harbin, China
[2] Faculty of Information Technology, Monash University, Australia
[3] Alibaba-NTU Joint Research Institute and also School of Computer Science and Engineering, NTU, Singapore
[4] Energy Research Institute, Nanyang Technological University, Singapore
[5] College of Intelligence and Computing, Tianjin University, Tianjin, China
[6] School of Computer Science and Engineering, NTU, Singapore
* The corresponding author, email: kavinkang@ntu.edu.sg

Abstract: As the 5G communication networks are being widely deployed worldwide, both industry and academia have started to move beyond 5G and explore 6G communications. It is generally believed that 6G will be established on ubiquitous Artificial Intelligence (AI) to achieve data-driven Machine Learning (ML) solutions in heterogeneous and massive-scale networks. However, traditional ML techniques require centralized data collection and processing by a central server, which is becoming a bottleneck of large-scale implementation in daily life due to significantly increasing privacy concerns. Federated learning, as an emerging distributed AI approach with privacy preservation nature, is particularly attractive for various wireless applications, especially being treated as one of the vital solutions to achieve ubiquitous AI in 6G. In this article, we first introduce the integration of 6G and federated learning and provide potential federated learning applications for 6G. We then describe key technical challenges, the corresponding federated learning methods, and open problems for future research on federated learning in the context of 6G communications.

Keywords: 6G communication; federated learning; security and privacy protection

I. INTRODUCTION

The rapid development of wireless communication techniques with numerous technological innovations for decades has greatly improved people's lives and promoting the development of the industry. As shown in Figure 1, the current Fourth-generation (4G) LTE network has created an era of mobile internet with Web search services, multimedia services, and APPs as core functions [1]. The upcoming Fifthgeneration (5G) system is designed to support a wider range of services, such as Augmented Reality/Virtual Reality (AR/VR), large-scale Internet of Things (IoT), and autonomous driving [2]. Specifically, the 5G system includes three technical characteristics: enhanced Mobile BroadBand (eMBB), massive Machine-Type-Communications (mMTC), ultra-Reliable LowLatency Communications (uRLLC). Since 5G brings unprecedented benefits to humans and is being actively deployed around the world, both industry and academia have begun to move towards the next generation of wireless technology, i.e., Sixth-generation (6G) [1], [2].

The 5G system represents a new wireless communication paradigm that adopts a ser-

Received: Jun. 4, 2019
Revised: Jul. 13, 2020
Editor: Ying-Chang Liang

In this article, we provided an overview of integrating federated learning into 6G communications.

vice-based architecture (SBA) instead of a communication-oriented architecture (COA) to achieve "connected things". In contrast to previous generations, 6G with transformative technologies will revolutionize the development of wireless communication from "connected things" to "connected intelligence" [1]. Specifically, 6G will revolutionize technology in three areas: new media, new services, and new infrastructures. It is expected that the 6G system will adopt advanced artificial intelligence (AI) technologies in these fields, and promptly and efficiently collect, transmit, and learn data anytime, anywhere to generate a large number of innovative applications and intelligent services [3]. In particular, ubiquitous AI will empower the promising 6G, a hyper-flexible architecture that brings human-centric development concepts to all aspects of network systems, instead of data-centric, machine-centric, and application-centric [4]. Therefore, 6G communications have higher-level security and stronger privacy protection requirements.

However, traditional Machine Learning (ML) empowered frameworks based on a central server are suffering from critical privacy and security challenges, e.g., single point of failure, which is not able to enable ubiquitous and secure AI for 6G. Moreover, due to large overhead caused by centralized data aggregation and processing, traditional centralized ML schemes might not be suitable for ubiquitous ML [5]. Thereby, decentralized ML solutions, in which all private data is kept in training devices locally, are becoming increasingly essential for 6G. Recently, Federated Learning (FL) as an emerging decentralized ML solution has attracted particular attention from academia and industry [2], [6]. In FL, participating devices collaboratively train a shared model through their local data, and thus only upload model updates instead of raw data to centralized parameter servers [7].

Although FL brings high potential for AI-empowered 6G and significantly improves privacy-sensitive applications with 6G communication, FL is still in the early stages of development and is facing new challenges in 6G scenarios. In this paper, we first introduce the core challenges of FL in 6G communications including (i) large communication cost due to multiple communication rounds for model updates and aggregation [8]; (ii) security problems caused by heterogeneous and diverse participating entities, e.g., poisoning attacks and backdoor attacks [7], [9]; (iii) privacy problems resulted from gradient leakage attacks and membership inference attacks [10]; (iv) model training and inference efficiency problems among massive-scale 6G networks. We then propose advanced federated learning methods to address the above challenges from different perspectives. Finally, we describe the open research topics and future directions of FL in 6G communications.

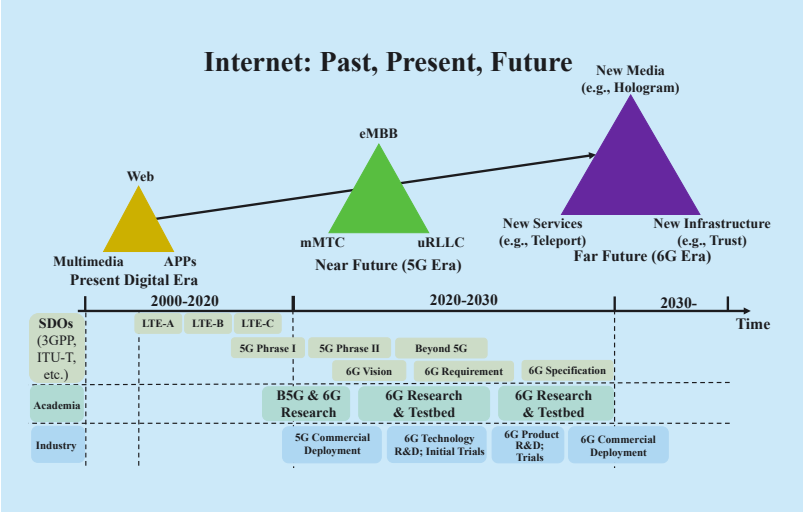


Fig. 1. *Key services and roadmap for 6G [1]–[4].*

II. Preliminaries and Overview

2.1 Key 6G requirements and use cases

Unlike 5G communications, 6G has prominent features to ensure ubiquitous, seamless, intelligent, high-performance connectivity and networking with security and privacy protection. More specifically, we will introduce the performance requirements of 6G communications as follows.

1) ***High Performance Networking****:* It is

commonly believed that 6G is a complex networking system with many heterogeneous space-air-ground-underwater communication networks [2], [4]. The three-dimensional super-connectivity networks provide worldwide connectivity and integrated networking to enable different types of network services and dense coverage through sub-networks and sub-systems, e.g., satellite communication networks, underwater-land communications. With the help of massive-scale heterogeneous networks, 6G communications can achieve up to 1 Tbps data rate per user, ultra-low end-to-end delay, superior end-to-end reliability, and high energy efficiency networking [2]. Compared with 5G communications, 6G communications support networking and connecting the majority not only in dense areas but also the less dense areas, such as the underwater environment, in an efficient and low overhead manner [4]. 6G communications employ novel communication networks to support highly diversified data, e.g., audio, video, AR/VR data, which reaches new communication experience with virtual networking existence and involvement anywhere [2].

2) ***Higher Energy Efficiency****:* In the 6G era, there exist higher energy efficiency requirements for wireless devices with charging constraints and battery life limitations. Therefore, long battery life and low energy consumption are two popular research topics for 6G communications. To address the energy problems of wireless devices, especially smartphones, existing studies have proposed energy harvesting technology, wireless power transfer technology, and green communication to improve energy efficiency and extend the working time of wireless devices [4].

Especially, the wireless devices can harvest energy from ambient radio-frequency, solar, geothermal energy, and wind energy by using different energy harvesting technologies, which can prolong the battery life. Similarly, the wireless devices with wireless charging equipment can obtain energy supplement from dense network infrastructures or mobile charging stations, e.g., Unmanned Aerial Vehicle (UAV), Electric Vehicles (EVs), through wireless power transfer technology.

Recently, to address energy problems for wireless devices, and emerging technology named symbiotic radio (SR) is introduced to integrate passive backscatter devices with an active transmission system [11], [12]. A typical example of SR is ambient backscatter communication, that enables network devices to utilize ambient RF signals to transmit information without requiring active RF transmission, making battery-free communication possible [11], [13]. Smart energy management is another promising mechanism with the goal of dynamically optimizing the balance between energy demand and supply [11].

For green communication techniques, AI-based solutions are quite important to optimize energy usage and energy scheduling for wireless devices in a dynamic environment and complex optimization goals. Advanced machine learning techniques, such as deep reinforcement learning, can be utilized to optimize the computation task offloading decision of a wireless device, and also make the best scheduling solution of working and sleeping time, which can lower energy consumption and enhance energy efficiency. The AI-based solutions can be also applied in multi-hop information routing in cooperative relay communication and communication infrastructure deployment in network-densification 6G scenarios, which significantly reduces the transmit power of the wireless devices without long propagation distance thus enabling high-efficiency communication [2], [4].

3) ***High Security and Privacy****:* Existing research mainly focuses on network throughput, reliability, and delay in 4G and 5G communications [4]. However, in the past few decades, wireless communication security and privacy issues have been ignored to some extent. Since data security and privacy issues are closely related to users' lives, protecting data security and privacy has become a very important part of human-centric 6G communications. Meanwhile, communication/data service providers legally collect a large amount of user infor-

mation, which leads to frequent leakage of privacy data. In order to solve this problem, it is envisaged that FL techniuqes can be used to achieve privacy-enhanced deep learning in 6G networks.

4) ***High Intelligence****:* The high intelligence of 6G will be beneficial to provide users with high-quality, personalized, and intelligent services. High-intelligent 6G includes operational intelligence, application intelligence, and service intelligence as follows.

- **Operational Intelligence.** Traditional network operations involve a series of resource optimization and multiobjective performance optimization problems [1]. In order to achieve a satisfactory level of network operation, optimization methods based on game theory, contract theory, etc. are widely used. However, these optimization theories may not obtain the optimal solution in large-scale timevarying variables and multi-objective scenarios. With the development of deep learning technologies, the above can be solved by using advanced machine learning technologies. On the other hand, the emergence of federated learning has transformed the multi-objective linear optimization problems into a nonlinear optimization problem, thus finding out the best solution for complex and timevarying decisions in operational intelligence [2].
- **Application Intelligence.** At present, applications related to 5G networks are gradually becoming intelligent. For 6G networks, intelligent applications are one baseline of application requirements [9], [14]. FL empowered wireless communication technologies to enable devices to connect with 6G networks to run a variety of intelligent applications. For example, in the future, users may need intelligent voice assistants to complete their daily schedules [15]. The 6G network ubiquitous AI will provide users with highly intelligent applications.
- **Service Intelligence** Furthermore, as a human-centric network, the high intelligence of the 6G network will provide intelligent services in a satisfactory and personalized manner [1], [2], [4]. For example, FL provides users with personalized healthcare services, personalized recommendation services, and personalized intelligent voice services in a distributed learning manner. In the future, intelligent services will be tightly integrated with the 6G networks [4].

5) ***Increased Device Density****:* Compared with 5G, the 6G has much higher transmission rates and shorter delay, greater device density, and the integration of Artificial Intelligence (AI). With the increased device density and explosive increasing data traffic, it is more and more important to solve the network capacity challenges. One of the potential solutions is to provide increasingly more but smaller radio cells that can transmit data quickly and energy-efficiently. These cells are required to be connected as seamlessly as possible to the fiber-optic core networks via high-performance transmission links. An important goal is to connect these wireless transmission links directly to fiber-optic networks without complex electronics. Thus the fiber optic networks can provide extremely high transmission capacity and reliability for massive devices with insignificant latency through flexible and ubiquitous wireless networks [16].

6) ***Green Communication****:* It is significant for green communication to make good decisions for optimizing resource utilization and communication efficiency. In 6G communication scenarios, due to massive network traffic, innumerable network devices, and dynamic network environments, there exist more and more complex resource optimization problems, e.g., green communication optimization and offloading decision, that traditional mathematical programming techniques and optimization solutions cannot tackle. Recently, data science and AI-based optimization have also largely been used to solve problems related to resource optimization, task assignment in distributed systems, because of its advantages of data-driven decision, dynamic flexibility, and self-adjustment.

2.2 Typical use scenarios

Compared with previous generations, the 5G service model has been transformed into a service-based architecture, and its user cases include: enhanced Mobile Broad Band (eMBB), massive Machine-Type-Communications (mMTC), and ultraReliable Low-Latency Communications (uRLLC). As shown in Figure 2, driven by Industry 5.0 and deep learning technologies, 6G will provide the following new service types:

- **New Media.** With the rapid development of wireless network communication technologies, it can be expected that the form of information interaction will gradually evolve from AR/VR to high fidelity extended reality (XR) interaction after 10 years, and even realizing wireless holographic communication [1]–[4]. Users can enjoy the new services brought by holographic communication and holographic display anytime and anywhere, such as virtual education, virtual tourism, virtual sports, virtual painting, virtual concerts, and other fully immersive holographic experiences.
- **New Services.** According to ITU-T's 6G communication technology white paper, beyond and high-precision teleport technology will provide users with a variety of new services [1]–[4], [17]. Holographic teleport, quantum communication, visible light communication (VLC), and other communication technologies have subverted the traditional service model. For example, industries such as remote surgery [18], cloud PLC [19], and intelligent transportation systems [20] will be empowered by the new service model to provide users with better services. The goal of these new technologies is to provide high-precision services, deterministic service, and bestguaranteed services.
- **New Infrastructure.** With the development of deep learning technologies, the 6G communication system has spawned many emerging infrastructures such as Integrated Terrestrial and Space [19], federated learning networks [14], decentralized infrastructures [1], and trustable infrastructure [4]. In particular, the FL network benefits from the high bandwidth and low latency of the 6G network, which has brought many emerging intelligent applications to cities, factories, and people.

2.3 Federated learning

In this subsection, we introduce an FL-based distributed learning architecture in 6G. In this architecture, a large number of decentralized devices associated with different services can collaboratively train a shared global model (e.g., anomaly detection, recommendation system, next-word prediction, etc.) by using locally collected datasets.

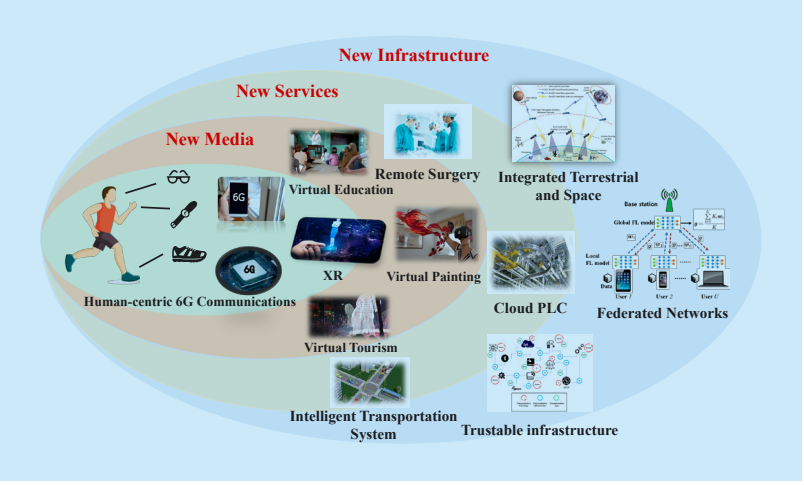


Fig. 2. *Key services and roadmap for 6G [21]–[23].*

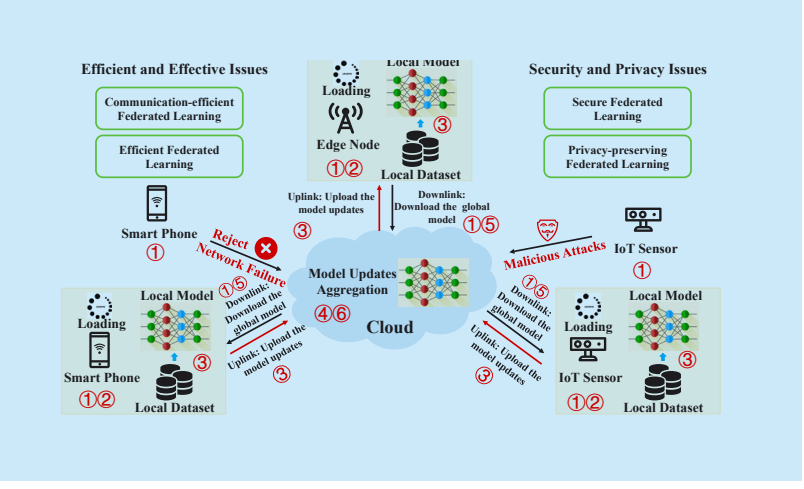


Fig. 3. *An overview of federated learning process in 6G [24].*

As shown in Figure 3, the procedure of FL-based architecture is divided into three phases: the initialization, the training, and the aggregation phase. In the initialization phase, a device will evaluate its service requests, needs, and connection conditions, and decides whether to register with the nearest cloud to join the training of the shared global model via a wired or wireless connection (e.g., 6G). Then, the cloud acting as task publisher will randomly select a subset of devices from the registered devices to participate in this round of training, and reject the remaining registered devices. The cloud will also send initialized or pre-trained global model ω_t to each selected device (steps ①, ②). In the training phase, each selected device trains global model $\omega_t^k \leftarrow \omega_t$ by by using local dataset to obtain the updated global model ω_{t+1}^k in each iteration. In particular, for the k-th device ($k \in \{1,2,\cdots,K\}$), the loss function needs to be optimized as follows: $\arg\min_{\omega \in R} F_k(\omega), F_k(\omega) = \frac{1}{D_k}\sum_{i \in D_k} f_i(\omega)$, where D_k denotes the size of local dataset that contains input-output vector pairs (x_i, y_i), $x_i, y_i \in R$, ω is local model parameter, and $f_i(\omega)$ is a local loss function (e.g., $f_i(\omega) = \frac{1}{2}(x_i^T \omega - y_i)$). Each selected device uploads the model updates to the cloud (steps ③, ④, ⑤). In the aggregation phase, the cloud receives model updates of all selected devices for aggregation to obtain a new global model ω_{t+1} for the next iteration, i.e., $\omega_{t+1} \leftarrow \omega_t - \frac{1}{K}\sum_{k=1}^{K} F_k(\omega)$, where K denotes the number of edge nodes. In the next round, the device selected by the cloud downloads the current latest global model ω_{t+1} from the cloud. The device will use the received new global model to update its respective model. In the next round of training, the cloud will randomly select a new device subset and repeat the above process until the trained model converges or meets the stopping criteria (step ⑥).

III. Core Challenges for Federated Learning in 6G

In this section, we introduce the core challenges of FL, which are the main bottleneck problem before large-scale deployment of FL in 6G applications.

3.1 Challenge 1: expensive communication

Since the FL involves thousands of devices participating during model training, communication is a critical bottleneck ffor FL being widely used in 6G [14]. Previous studies [6], [8], [10], [24]–[28] has made many efforts to improve the communication efficiency of FL system. Furthermore, it is challenging for FL networks to achieve communication in the FL networks is synchronized with the local calculation of the device [8], [10], [29]. To make the FL model suitable for 6G networks with massive, heterogeneous devices and networks, it is necessary to develop a communication-efficient method, which can greatly reduce the number of gradients exchanged between the devices and the cloud instead of all gradients information. In order to further reduce communication overhead in this setting, two key aspects need to be considered: (i) reducing the total number of communication rounds, or (ii) reducing the number of gradients in each communication round.

3.2 Challenge 2: security problems

Since 6G networks can provide ubiquitous services across a wider geographic area, the computing and communication capabilities of each device in the network may vary due to changes in hardware (CPU, GPU), network connectivity (4G, 5G, 6G, WiFi), and energy (battery level). Obviously, the system heterogeneity between the devices will bring some confusion and faults to the FL model and 6G network [9], [14]. Additionally, there may be unreliable devices in the FL, which may cause the Byzantine failure of the system. Similarly, adversaries may launch active learning-based attacks (like poisoning attacks and backdoor

attacks) on heterogeneous devices and cause errors in the FL system. The security vulnerabilities of these FL systems greatly exacerbate challenges such as mitigating attacks, tolerance, and faults. Therefore, developing a secure and robust FL must: (i) defend against malicious attacks, (ii) tolerate heterogeneous hardware, and (iii) achieve robust aggregation algorithms.

3.3 Challenge 3: privacy concerns

Although FL protects the privacy of each device by sharing model updates (e.g., gradients information) instead of the raw data, the private data will still be disclosed during the interaction between the device and the cloud [30]. For example, adversaries will launch **membership inference** [30] or **gradient leakage attacks** [31] to steal local training data from the devices. Previous work has focused on using tools such as secure multi-party (SMC) computing or homomorphic encryption (HE) to enhance the privacy of FL, but these methods cannot address the above malicious attacks [14]. SMC and HE can only prevent data breach problems and cannot resist member inference attacks and gradient leak attacks. Therefore, it is very urgent for the FL system to develop new privacy-enhancing techniques to resist or mitigate the aforementioned malicious attacks.

3.4 Challenge 4: effective issues

Deploying FL models to devices generally involves model training and inference [14]. If the speed of model training and reasoning is relatively slow, users will not be able to experience real-time intelligent services [32]. Therefore, when FL systems are widely deployed in 6G networks, they will encounter the following challenges: (i) the size of the FL model is too large to adapt to a single device; (ii) the FL model training is too slow to meet the delay requirements of the 6G network; (iii) the FL model inference is too slow to satisfy the user' real-time demand. Efficient training and inference are necessary for the perfect integration of FL and 6G networks. However, it is challenging for FL systems to achieve efficient model training and inference in a massive, heterogeneous network.

IV. Advanced Federated Learning Methods for 6G

To address the aforementioned challenges, we propose advanced federated learning systems through different emerging technologies or methods to enable communication-efficient, secure, and privacy-enhanced federated learning, respectively.

4.1 Communication-efficient federated learning for 6G

In 6G, it is challenging for devices that span a larger geographic area to obtain a better global model but with huge communication overhead. The communication overhead will affect the gradient exchange between the devices and the cloud, thus affecting the model aggregation at the cloud. Therefore, we need to find a more efficient way to achieve FL training. In this subsection, we will explain communicationefficient FL from the perspectives of system-level and algorithm level, which promotes a wider-range FL deployment and usage for 6G communications.

*1) **Communication-efficient FL: System Level**:* From a system perspective, data distribution (e.g., non-independent and identical distribution), device distribution (e.g., heterogeneous devices across regions and networks),

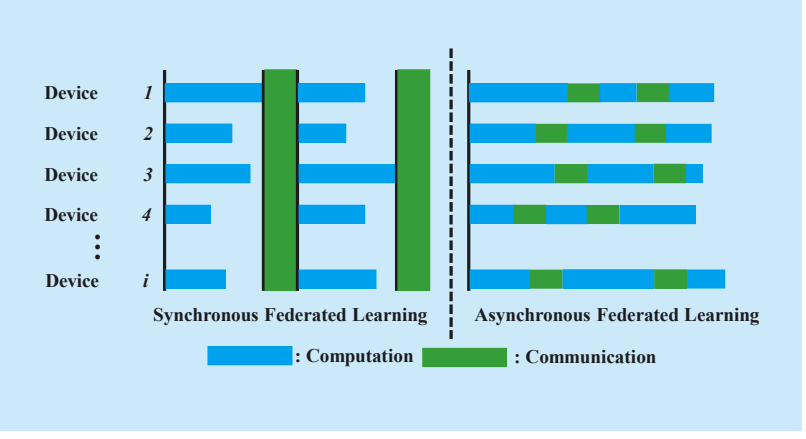


Fig. 4. *The overview of the synchronous and asynchronous FL.*

computation methods (e.g., decentralized and centralized), and communication mechanisms (e.g., synchronous and asynchronous scheme) have different impacts on communication efficiency in different application scenarios [8].

- **Asynchronous FL System:** As shown in Figure 4, AFLS can reduce the computation time of the devices by asynchronously aggregating the model updates, thereby improving the communication efficiency of FL. Let $\kappa = \frac{Comm}{Comp + Comm}$, where κ represents communication efficiency, Comm is the communication time, and Comp is the computation time. It can be seen from Figure 4 that the Comp of asynchronous model update scheme is shorter than that of synchronous one, so the communication efficiency κ of the asynchronous model update scheme is higher than that of the synchronous one.

*2) **Communication-efficient FL: Algorithm Level**:* At the algorithm level, achieving communication-efficient FL can reduce the communication rounds of training a model by accelerating convergence [29] and reduce the communication cost of each round by using gradient compression techniques [10](e.g., sparsification, quantization, etc.). More details are described below.

- **Accelerating Model Convergence:** Stochastic gradient descent (SGD) algorithms based on zero-order, first-order, second-order, and federated optimization are used to reduce the number of rounds of model training [8]. Since the federated optimization method can protect the private data on each device, it is very popular with this unique motivation in accelerating training model convergence.
- **Reducing Communication Overhead:** Gradient sparsification and gradient quantization can greatly reduce the large number of gradients exchanged between the devices and the cloud to achieve communication-efficient FL. Lin *et al.* in [10] proposed a Top-k selection-based gradient compression scheme to improve communication efficiency. In this scheme, the authors can compress the gradient 300 times to reduce the number of gradients without compromising accuracy.

4.2 Secure federated learning for 6G

Due to the wide range of 6G network connections, FL will suffer malicious attacks from heterogeneous networks, heterogeneous devices, and malicious participants during the training process [9]. To alleviate this problem, researchers have proposed many different defense solutions from three perspectives: aggregation algorithm, detection mechanism, and reputation management.

*1) **Robust Aggregation Algorithm**:* Aggregation is a very important operation in the FL training process that directly affects the results of model convergence. The motivation of the robust aggregation algorithm is to greatly reduce the impact of low-quality model updates generated by malicious devices (i.e., poisoning attacks) on global model training. Furthermore, this method can make the cloud tolerate Byzantine failures of some devices [33], [34]. For example, Ang *et al.* in [35] proposed the regularizer approximation method to reduce the noise interference of heterogeneous devices and heterogeneous networks.

*2) **Robust Detection Mechanism**:* Another intuitive idea is to detect malicious devices to prevent them from participating in FL training. Such a mechanism has generally utilized the accuracy of the sub-model generated by the device as an evaluation metric to detect malicious devices. Liu *et al.* in [9] utilized the smart contract techniques in the blockchain to design a malicious device detection mechanism to alleviate the malicious attack problems.

*3) **Reliable Reputation Management**:* The historical behaviors of the devices can be used as a key indicator to evaluate its reliability and trustworthiness by a metric named reputation. The high reputation value indicates more reliable devices. Inspired by this, establishing a reputation management scheme for device historical behaviors in FL can also prevent

malicious devices from damaging the global model. Kang *et al.* in [15] proposed a reputation management scheme to calculate the historical reputation of the devices to achieve a robust FL with high-reputation devices.

4.3 Privacy-preserving federated learning for 6G

1) ***Differentially Privacy****:* Differential privacy (DP) [36] techniques are proposed to protect the privacy of gradient information, thereby achieving cloud-level privacy protection. Geyer *et al.* in [37] applied the DP technique in FL system that protects cloud-level privacy. Similarly, in order to protect userlevel privacy, the local differential privacy (LDP) techniques achieve this goal by disturbing the gradients uploaded by the devices [30]. However, DP and LDP technologies enhance FL privacy at the expense of model performance. Therefore, there are currently advanced methods that balance privacy and performance as described below.

2) ***Deep Net Pruning****:* Neural network pruning is a technique of deep learning whose goal is to develop a smaller and more efficient neural network. Recently, Huang *et al.* [38] utilized pruning as an equivalent technique of DP to protect the privacy of the FL system while ensuring the model performance. Such a method creates a new idea of using model pruning to be equivalent to DP techniques, which provides new opportunities for balancing utility and privacy.

3) ***Gradient Compression****:* The reason why adversaries can infer the local data of the devices is that the gradient information contains rich semantic information [31]. Inspired by the above, an intuitive idea is that the methods that disrupt the distribution of gradient information thus protecting the gradient privacy. Zhu *et al.* in [31] proved that gradient compression can defend against gradient leakage attacks without compromising accuracy and the defense effect is better than that of DP.

4.4 Effective federated learning for 6G

The long-term goal of human-centric communication services in 6G networks is to handle user needs in real time. Therefore, it is necessary to achieve efficient FL from training and inference.

1) ***Efficient Training****:* Efficient training can greatly reduce the training time of mobile devices to achieve efficient FL. The advanced training methods are summarized as follows.

- **Federated Parallelization:** Data and model parallelization are generally used to accelerate model training. Data parallization achieves efficient training by running multiple training samples in parallel [39], [40]. Model parallelization accelerates model training by splitting the model over multiple processors [32].
- **Federated Distillation:** Model distillation adopts transfer learning to utilize the output of a pre-trained complex model (i.e., Teacher model) as a supervised signal to train another simple network, i.e., Student model. Such a way can train student models to improve the efficiency of model training. Jeong *et al.* in [26] proposed federated distillation (FD), an efficient distributed model training algorithm, whose training efficiency is much smaller than the FL benchmark scheme, especially when the model size is large.

2) ***Efficient Inference:****:* The size of the existing FL model is too large to realize real-time inference on the devices. Efficient inference can be achieved in the following ways.

- **Pruning:** The pruning technique is a model optimization technique that includes removing excess weights in the weight tensor. The compressed neural network not only runs faster but also reduces the computational cost of the training network, which is a critical step in deploying the model to mobile phones or other edge devices.
- **Weight Sharing.** Weight sharing reduces the number of model parameters by sharing weights, thereby achieving efficient mod-

el inference. The reason is that the fewer the parameters of the model, the smaller the model size. Tran *et al.* in [27] utilized weight sharing approach for wireless networks to improve model inference efficiency.

V. Open Research Topics and Future Directions

5.1 Trustworthy federated learning

1) ***Privacy-enhanced Federated Learning****:* Previous work about FL has covered user or cloud-level privacy for all devices in the 6G networks. However, in practice, the previous schemes provide strict privacy restrictions at the expense of accuracy [14]. It is essential for FL to develop privacyenhanced techniques that do not compromise accuracy to provide strict privacy guarantees because the industry is very concerned about the accuracy of the FL model. To this end, few studies are exploring potential solutions. For example, Huang *et al.* [38] recently proposed a net pruning technique to provide strict privacy guarantees by replacing pruning with DP technique, and also to improve the training efficiency of the model. It is an interesting and ongoing direction to developing methods that can balance efficiency and privacy restrictions in future work.

2) ***Security-enhanced Federated Learning****:* Since the FL systems normally involve multiple entities of devices, cloud, and machine learning model providers, it is vulnerable to malicious attacks from adversaries against different entities. Although existing work has made a lot of efforts to provide strong security protection for the FL systems, there is little work to defend or mitigate these malicious attacks from the system perspective. Bonawitz *et al.* in [28] explored several more secure and robust aggregation algorithms and fault tolerance mechanisms from the perspective of system design. The serurity-enhanced techniques are designed from a system perspective so that FL can develop more practical industrial applications with the help of 6G networks.

3) ***Fair Federated Learning****:* FL involves thousands of devices training a shared global model in massive, heterogeneous networks [41]. Naive optimizing the global model in such a network may be unfair to some devices by causing disproportionate advantages or disadvantages. Obviously, FL towards fairness is an indispensable requirement for humancentric 6G communication services. Specifically, a fair FL in a wireless network involves fair resource allocation and a reasonable incentive mechanism. How to allocate computing and communication resources accurately and fairly in massive, heterogeneous networks has become a critical challenge that needs to be solved urgently. Some pioneering work, Li *et al.* in [41] proposed q-Fair FL (q-FFL), which is a new aggregation algorithm to achieve a fair allocation of resources and accuracy.

4) ***Explainable Federated Learning****:* The vast majority of FL models are black-box models (i.e., without interpretability), which makes users unable to understand what kind of services the model provides for themselves. In a complex 6G network system, the unexplainable predictions or decisions output by the black-box model may cause huge losses to users. For example, 6G-supported self-driving relies on an on-vehicle visual recognition model to determine whether the vehicle is running or stopped. Since the on-vehicle model has no interpretability, the driver cannot understand the decision of the vehicle model output. In 2018, the self-driving vehicle developed by Uber caused a car accident due to the wrong output of the on-vehicle black-box model[1]. Therefore, in the context of a complex network system, such as 6G, the development of an interpretable FL model is the necessary way to human-centric communication services.

5.2 Efficient and effective federated learning

1) ***Novel Asynchronous System****:* Even though the 6G network can bring the advantage of extremely low latency to the FL systems, the communication overhead is still the bottleneck

[1] http://tech.sina.com.cn/zt˙d/uberincident/

of the FL systems being widely used [14]. As described in Section IV-A1, the two most commonly studied communication optimization schemes in distributed machine learning systems are the batch synchronous method and the asynchronous method (where the delay of the model update is assumed to be bounded) [28]. Indeed, asynchronous communication schemes involve scheduler, coordinator, worker, and updater, so there are several optimization problems for these roles that can be considered in the future: i) how the scheduler reasonably schedule the communication and computing resources in the systems; ii) how the coordinator efficiently control the working state and idle state of the devices; iii) how workers and updaters optimize hyperparameters for model updates. These optimization problems are worth studying in future work in order to develop novel asynchronous systems.

*2) **Neural Architecture Search**:* The structure of the current FL models is generally predefined, but this predefined architecture may not be the best choice because it may not be suitable for non-independent and identical distribution (nonIID) data. Therefore, the Neural Architecture Search (NAS)based Automating FL (AutoFL) schemes may be a promising solution to this problem. For example, a study in [42] proposed a federated NAS (FedNAS) algorithm to help distributed devices collaborate to find a better architecture with higher accuracy. NAS provides opportunities for seeking a better FL model architecture in the future.

5.3 Towards incentive federated learning

Existing studies mainly focus on enhancing the performance of FL algorithms, e.g., accuracy and training time. Nevertheless, an optimistic assumption, that all the data owners are willing to join the FL anytime and anywhere, is not practical in 6G scenarios with massive self-interest devices. As a result, incentive mechanisms for honest and active participation are a core and urgent research topic [7], [25], [43]–[46]. Some interesting topics include: i) Due to information asymmetric between task publishers and participating devices, e.g., information about time-varying available resources, unfixed working periods and changeable participation willingness, it is still an open issue to design effective online-learning based incentive mechanisms to remove the impacts of both information asymmetric and time-varying factors, and also ensure efficient federated learning in 6G scenarios; ii) Considering heterogeneous and massive devices with diverse hardware equipment in 6G scenarios, the data quality of the devices is diverse. But the data quality plays an important role in learning performance. It is a challenging problem how to design data quality-based incentive mechanisms to motivate more devices with high-quality data to participate in federated learning and obtain higher rewards for their high-quality data contributions, thus improving both the system reliability and the learning performance [7], [43], [45].

5.4 Towards personalized federated learning

It is challenging for the FL system in the 6G network to provide users with personalized services. Prior studies [47]–[50] adopt different personalized techniques to provide users with real-time personalized services, which is a solid step towards personalized FL. However, personalized FL still faces challenges from non-IID data, system heterogeneity, and network heterogeneity. Personalized service is a very important part of the human-centric 6G services. Therefore, it is an interesting and meaningful topic for FL to seek novel ways to address the above challenges.

VI. Conclusion

In this article, we provided an overview of integrating federated learning into 6G communications. We discussed the requirements of 6G communication and core challenges of federated learning for 6G applications. For the above challenges, we provided a comprehensive introduction of the emerging advanced federated

learning methods for 6G communications, which including communication-efficient federated learning, secure federated learning, and effective federated learning. Finally, we outlined out a handful of open problems and directions worth future research efforts.

Acknowledgement

This research is supported by the National Research Foundation (NRF), Singapore, under Singapore Energy Market Authority (EMA), Energy Resilience, NRF2017EWT-EP003-041, Singapore NRF2015-NRF-ISF001-2277, Singapore NRF National Satellite of Excellence, Design Science and Technology for Secure Critical Infrastructure NSoE DeST-SCI2019-0007, A*STARN-TU-SUTD Joint Research Grant on Artificial Intelligence for the Future of Manufacturing RGANS1906, Wallenberg AI, Autonomous Systems and Software Program and Nanyang Technological University (WASP/NTU) under grant M4082187 (4080), and NTU-WeBank JRI (NWJ-2020-004), Alibaba Group through Alibaba Innovative Research (AIR) Program and Alibaba-NTU Singapore Joint Research Institute (JRI), NTU, Singapore, and National Key Research and Development Program of China under Grant 2018YFC0809803 and Grant 2019YFB2101901, Young Innovation Talents Project in Higher Education of Guangdong Province, China under grant No. 2018KQNCX333, in part by the National Science Foundation of China under Grant 61702364.

References

[1] K. B. Letaief, W. Chen *et al.*, "The roadmap to 6g: Ai empowered wireless networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[2] Y. Xiao, G. Shi, and M. Krunz, "Towards ubiquitous ai in 6g with federated learning," *arXiv preprint arXiv:2004.13563*, 2020.

[3] K. David and H. Berndt, "6g vision and requirements: Is there any need for beyond 5g?" *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80, 2018.

[4] S. Dang, O. Amin *et al.*, "What should 6g be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[5] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities and challenges," *arXiv preprint arXiv:1908.06847*, 2019.

[6] J. Konec˘ny`, H. B. McMahan *et al.*,"Federated learning: Strategies for improving communication efficiency," *arXiv preprint arXiv:1610.05492*, 2016.

[7] J. Kang, Z. Xiong *et al.*, "Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 700–10 714, 2019.

[8] Y. Shi, K. Yang *et al.*, "Communication-efficient edge ai: Algorithms and systems," *arXiv preprint arXiv:2002.09668*, 2020.

[9] Y. Liu, J. Peng *et al.*, "A secure federated learning framework for 5g networks," *arXiv preprint arXiv:2005.05752*, 2020.

[10] Y. Lin, S. Han *et al.*, "Deep gradient compression: Reducing the communication bandwidth for distributed training," in *International Conference on Learning Representations*, 2018. [Online]. Available: https://openreview.net/forum?id=SkhQHMW0W

[11] T. Huang, W. Yang *et al.*, "A survey on green 6g network: Architecture and technologies," *IEEE Access*, vol. 7, pp. 175 758–175 768, 2019.

[12] R. Long, H. Guo *et al.*,"Full-duplex backscatter communications in symbiotic radio systems," *IEEE Access*, vol. 7, pp. 21 597–21 608, 2019.

[13] G. Yang, Q. Zhang *et al.*, "Cooperative ambient backscatter communications for green internet-of-things," *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1116–1130, 2018.

[14] T. Li, A. K. Sahu *et al.*, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

[15] J. Kang *et al.*, "Reliable federated learning for mobile networks," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 72–80, 2020.

[16] M. Giordani *et al.*,"Toward 6g networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[17] Z. Zhang, Y. Xiao *et al.*, "6g wireless networks: Vision, requirements, architecture, and key technologies," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 28–41, 2019.

[18] S. Nayak and R. Patgiri, "6g communication technology: A vision on intelligent healthcare," *arXiv preprint arXiv:2005.07532*, 2020.

[19] M. Giordani, M. Polese *et al.*, "Toward 6g networks: Use cases and technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[20] Y. Liu *et al.*, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet of Things Journal*, pp. 1–1, 2020.

[21] S. Gu, J. Jiao *et al.*, "Arma-based adaptive cod-

ing transmission over millimeter-wave channel for integrated satellite-terrestrial networks," *IEEE Access*, vol. 6, pp. 21 635–21 645, 2018.

[22] M. Chen, Z. Yang *et al.*, "A joint learning and communications framework for federated learning over wireless networks," *arXiv preprint arXiv:1909.07972*, 2019.

[23] A. Souri, A. Hussien *et al.*, "A systematic review of iot communication strategies for an efficient smart environment," *Transactions on Emerging Telecommunications Technologies*, p. e3736, 2019.

[24] L. U. Khan, N. H. Tran *et al.*, "Federated learning for edge networks: Resource optimization and incentive mechanism," *arXiv preprint arXiv:1911.05642*, 2019.

[25] J. Kang *et al.*, "Incentive design for efficient federated learning in mobile networks: A contract theory approach," in *2019 IEEE VTS Asia Pacific Wireless Communications Symposium*. IEEE, 2019, pp. 1–5.

[26] E. Jeong, S. Oh *et al.*, "Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data," *arXiv preprint arXiv:1811.11479*, 2018.

[27] N. H. Tran, W. Bao *et al.*, "Federated learning over wireless networks: Optimization model design and analysis," in *IEEE Conference on Computer Communications*. IEEE, 2019, pp. 1387–1395.

[28] K. Bonawitz, H. Eichner *et al.*, "Towards federated learning at scale: System design," in *SysML 2019*, 2019, to appear. [Online]. Available: https://arxiv.org/abs/1902.01046

[29] A. F. Atiya and A. G. Parlos, "New results on recurrent network training: unifying the algorithms and accelerating convergence," *IEEE transactions on neural networks*, vol. 11, no. 3, pp. 697–709, 2000.

[30] Z. Wang, M. Song *et al.*, "Beyond inferring class representatives: Userlevel privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2512–2520.

[31] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Advances in Neural Information Processing Systems*, 2019, pp. 14 747–14 756.

[32] L. Li, H. Xiong *et al.*,"Smartpc: Hierarchical pace control in real-time federated learning system," in *2019 IEEE Real-Time Systems Symposium (RTSS)*. IEEE, 2019, pp. 406–418.

[33] A. Portnoy and D. Hendler, "Towards realistic byzantine-robust federated learning," *arXiv preprint arXiv:2004.04986*, 2020.

[34] S. Guo, T. Zhang *et al.*, "Towards byzantine-resilient learning in decentralized systems," *arXiv preprint arXiv:2002.08569*, 2020.

[35] F. Ang, L. Chen *et al.*, "Robust federated learning with noisy communication," *IEEE Transactions on Communications*, 2020.

[36] M. Abadi, A. Chu *et al.*, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 308–318.

[37] R. C. Geyer, T. Klein *et al.*, "Differentially private federated learning: A client level perspective," *arXiv preprint arXiv:1712.07557*, 2017.

[38] Y. Huang, Y. Su *et al.*, "Privacy-preserving learning via deep net pruning," *arXiv preprint arXiv:2003.01876*, 2020.

[39] T.-D. Cao, T. Truong-Huu *et al.*, "A federated learning framework for privacy-preserving and parallel training," *arXiv preprint arXiv:2001.09782*, 2020.

[40] Z. Jiang, A. Balu *et al.*, "Collaborative deep learning in fixed topology networks," in *Advances in Neural Information Processing Systems*, 2017, pp. 5904–5914.

[41] T. Li, M. Sanjabi *et al.*,"Fair resource allocation in federated learning," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=ByexElSYDr

[42] C. He, M. Annavaram *et al.*, "Fednas: Federated deep learning via neural architecture search," *arXiv preprint arXiv:2004.08546*, 2020.

[43] L. U. Khan, N. H. Tran *et al.*, "Federated learning for edge networks: Resource optimization and incentive mechanism," *arXiv preprint arXiv:1911.05642*, 2019.

[44] J. Weng, J. Weng *et al.*, "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Transactions on Dependable and Secure Computing*, 2019.

[45] Y. Zhan, P. Li *et al.*, "A learning-based incentive mechanism for federated learning," *IEEE Internet of Things Journal*, 2020.

[46] H. Yu, Z. Liu *et al.*, "A fairness-aware incentive scheme for federated learning," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2020, pp. 393–399.

[47] A. Fallah, A. Mokhtari *et al.*, "Personalized federated learning: A metalearning approach," *arXiv preprint arXiv:2002.07948*, 2020.

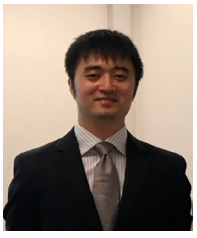
[48] Q. Wu, K. He *et al.*, "Personalized federated learning for intelligent iot applications: A cloud-edge based framework," *IEEE Open Journal of the Computer Society*, pp. 1–1, 2020.

[49] R. Hu, Y. Guo *et al.*, "Personalized federated learning with differential privacy," *IEEE Internet of Things Journal*, 2020.

[50] V. Kulkarni, M. Kulkarni *et al.*, "Survey of personalization techniques for federated learning," *arXiv preprint arXiv:2003.08673*, 2020.

Biographies

Yi Liu (S'19), received the B.Eng. degree in network engineering from Heilongjiang University, Harbin, China, in 2019. He is currently pursuing a Ph.D. degree at the Faculty of Information Technology, Monash University, Melbourne, Australia. His research interests include security & privacy, federated learning, edge computing, and blockchain.



Xingliang Yuan, obtained his PhD degree in Computer Science from City University of Hong Kong, China in 2016. Before that, he received his MS degree and BS degree from Illinois Institute of Technology and Nanjing University of Posts and Telecommunications, respectively, both majored in Electrical Engineering. He is currently a lecturer with the Faculty of Information Technology, Monash University, Australia. His research has been supported by CSIRO Data61, Oceania Cyber Security Centre, Monash Infrastructure, the Hong Kong Innovation and Technology Commission, Amazon Web Services, and Microsoft Azure. His research focuses on designing protocols and systems to address privacy and security issues in cloud and networked applications. In the past few years, his work has appeared in prestigious venues in security, computer networks, and distributed systems, such as ACM CCS, ACM AsiaCCS, ESORICS, IEEE INFOCOM, IEEE ICDCS, IEEE ICNP, IEEE ICDE, IEEE TDSC, IEEE TIFS, IEEE/ACM TON, IEEE TPDS, IEEE JSAC, IEEE TMC, etc.

Zehui Xiong (S'17), is currently a researcher with Alibaba-NTU Singapore Joint Research Institute, Singapore. He received the Ph.D. degree at Nanyang Technological University, Singapore. He was a visiting scholar at Princeton University and University of Waterloo. His research interests include blockchain and edge intelligence. He has won several Best Paper Awards including IEEE WCNC, and IEEE VTS Singapore Best Paper Award in 2019. He is an Editor for Elsevier Computer Networks. He serves as a Guest Editor for IEEE Transactions on Cognitive Communications and Networking. He is the recipient of the Chinese Government Award for Outstanding Students Abroad in 2019.

Jiawen Kang, received the M.S. degree from the Guangdong University of Technology, China, in 2015, and the Ph.D. degree at the same school in 2018. He is currently a postdoc at Nanyang Technological University, Singapore. His research interests mainly focus on blockchain, security and privacy protection in wireless communications and networking.

Xiaofei Wang (S'06, M'13, SM'18) is currently a Professor with the Tianjin Key Laboratory of Advanced Networking, School of Computer Science and Technology, Tianjin University, China. He got master and doctor degrees in Seoul National University from 2006 to 2013, and was a Post-Doctoral Fellow with The University of British Columbia from 2014 to 2016. Focusing on the research of socialaware cloud computing, cooperative cell caching, and mobile traffic offloading, he has authored over 100 technical papers in the IEEE JSAC, the IEEE TWC, the IEEE WIRELESS COMMUNICATIONS, the IEEE COMMUNICATIONS, the IEEE TMM, the IEEE INFOCOM, and the IEEE SECON. He was a recipient of the National Thousand Talents Plan (Youth) of China. He received the "Scholarship for Excellent Foreign Students in IT Field" by NIPA of South Korea from 2008 to 2011, the "Global Outstanding Chinese Ph.D. Student Award" by the Ministry of Education of China in 2012, and the Peiyang Scholar from Tianjin University. In 2017, he received the "Fred W. Ellersick Prize" from the IEEE Communication Society.

Dusit Niyato (M'09-SM'15-F'17), is currently a professor in the School of Computer Science and Engineering, at Nanyang Technological University, Singapore. He received B.Eng. from King Mongkuts Institute of Technology Ladkrabang (KMITL), Thailand in 1999 and Ph.D. in Electrical and Computer Engineering from the University of Manitoba, Canada in 2008. His research interests are in the area of energy harvesting for wireless communication, Internet of Things (IoT) and sensor networks.