Personalized Online Federated Learning for IoT/CPS: Challenges and Future Directions

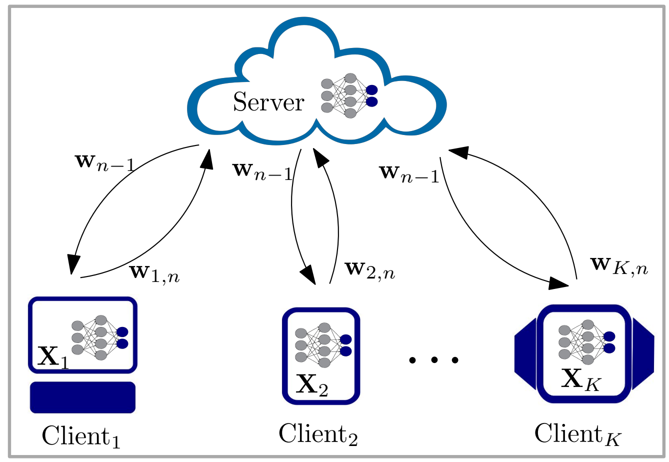
Vinay Chakravarthi Gogineni, *Member*, *IEEE*, Stefan Werner, *Senior Member*, *IEEE*, Francois Gauthier, *Member*, *IEEE*, Yih-Fang Huang, *Life Fellow*, *IEEE*, and Anthony Kuh, *Fellow*, *IEEE*

***Abstract*—In recent years, federated learning (FL) has emerged as a powerful paradigm for distributed learning thanks to its privacy-preserving capabilities. With the use of FL, a network of edge devices can make intelligent decisions without exposing their data to others. Despite its success, the traditional FL is not well suited to many practical applications such as those that involve the internet-of-things (IoT) or cyber-physical systems (CPS), where data access can be intermittent, and edge devices are semi-independent with device-specific dynamic behavior characteristics. Those devices are referred to here as *semi-independent* devices since they need to make decisions based on their own data and device characteristics, often independent of other devices and the information obtained from other devices in the network. Additionally, as new information becomes available, traditional FL must repeat the entire learning process and may not be able to provide timely and tailored solutions to participants. Personalized online FL, on the other hand, retains the collaborative and privacy-preserving aspects while learning in real time from intermittent data. It further enables devices to learn models customized to the device and the specific tasks it performs. In light of these reasons, personalized Online-FL is ideal for applications where the learning relies on heterogeneous data streams, and local optimization is beneficial. In this work, we want to bring attention to this new learning paradigm, present a few of the applications that could benefit from it, and highlight the principal challenges the research community faces in developing successful personalized Online-FL.**

***Index Terms*—Personalized online federated learning, internet-of-things, cyber-physical systems, distributed multitask learning.**

I. Introduction

Advances in telecommunication and semiconductor chip design technologies have contributed to the rapid development of the internet-of-things (IoT) and cyber-physical systems (CPS). Myriads of dispersed sensors and devices in these systems constantly gather data for inference and decision-making. Due to various factors (e.g., resource constraints of devices, channel capacity and availability, and privacy and data integrity of end users), it is usually not practical to transfer the data collected over edge devices to a central processing server or the cloud through wireless media. Considering these concerns, it is, therefore, imperative that large-scale networks establish a trustworthy and reliable distributed learning framework capable of learning from heterogeneous data arising over a myriad of devices with varying capabilities while complying with the individual privacy preferences of data holders. A new distributed learning framework referred to as federated learning (FL) [1] addresses many of the abovementioned issues associated with cloud-based centralized data processing.



$\mathbf{X}_k$: dataset at client k, $\mathbf{w}_{k,n}$: kth client local model, $\mathbf{w}_n$: global model

Fig. 1. Traditional federated learning.

An illustration of the traditional FL workflow is given in Fig. 1, in which numerous devices (e.g., IoT devices), referred to as clients, residing at the network edge perform local learning on their own fixed amount of data $\mathbf{X}_k$. These clients transfer their updated models $\mathbf{w}_{k,n}$ to an access point or server, where a global model $\mathbf{w}_n$ is developed by aggregating the local models. This aggregation could be a simple or weighted average of local models, depending on the need to emphasize specific clients' data. Afterward, the server communicates the aggregated global model back to the edge devices to replace their local model. This local learning-aggregation procedure continues for a predefined number of iteration rounds or until a pre-specified convergence criterion is met, allowing a network of devices to learn from each other. Since clients do not disclose their data during model training, FL ensures the data integrity of the clients contributing to the learning process. Additionally, since FL does not share the raw data, it can seamlessly handle unbalanced data across large-scale networks

The Research Council of Norway supported this work.

Vinay Chakravarthi Gogineni, Stefan Werner and Francois Gauthier are with the Department of Electronic Systems, Norwegian University of Science and Technology-NTNU, Trondheim 7491, Norway. Stefan Werner is also with the Department of Signal Processing and Acoustics, Aalto University, 00076, Finland

Yih-Fang Huang is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: huang@nd.edu).

Anthony Kuh is with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, HI 96822 USA (e-mail: kuh@hawaii.edu).
Anthony Kuh acknowledges support in part by NSF Grant 2142987.

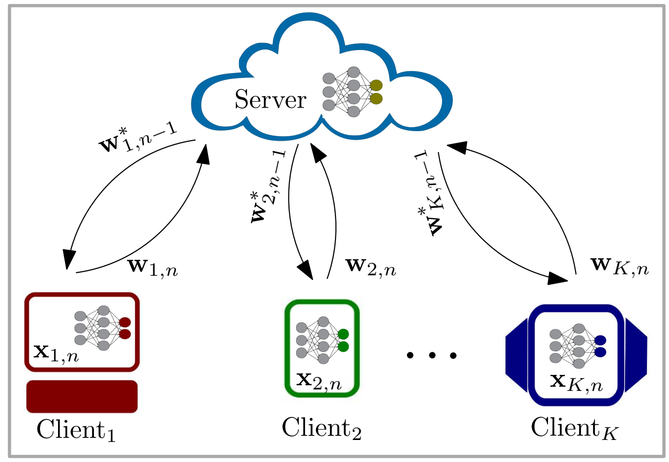
with varying statistical properties, such as IoT.

An IoT network is one of many applications in which participating clients receive data progressively or even a continuous stream of data. Data streams collected by edge devices in IoT networks may be used to perform local learning in real-time and to update global models. Importantly, the underlying model is more likely to evolve with time. In this regard, more significant weightage should be given to recent data during the learning process. Furthermore, in a few other applications, e.g., wireless communications and networked vehicles, the real-time responsiveness of the model is paramount. Because the traditional FL learns from fixed data batches that are not usually timely enough, it is not appropriate for these scenarios. Online federated learning (Online-FL) [2] is a viable solution for applications of this kind. In Online-FL, the clients perform local learning on heterogeneous data streams, then share their updated local models with the server to build a more accurate global model in real time. Thus Online-FL is better suited to learn from streaming data in a computationally efficient manner without requiring periodic retraining of the model.

Besides heterogeneous data streams, edge devices act semi-independently in the aforementioned applications and exhibit device-specific dynamic behavior. Clearly, this requirement highlights the importance of learning device-specific models rather than a single universal model. Therefore, this article focuses primarily on personalized Online-FL, which best fits the IoT/CPS featuring heterogeneous data streams. To this end, we provide an overview of personalized Online-FL and the applications it can serve. Furthermore, we present some challenges and potential future directions of personalized Online-FL in real-world settings to inspire new research that may bring the personalized federated IoT/CPS into reality. Our discussions will differ significantly from existing articles that deal primarily with simple traditional FL [3]–[5].

II. Personalized Online Federated Learning: Why and What Is It?

In numerous applications like IoT networks, IoT edge devices continually sense and collect data streams. These data streams may have different statistical properties due to their geographical dispersion or intrinsic characteristics of the underlying processes. The more critical aspect of these edge devices is that they behave semi-independently and collaborate among themselves to improve their decision-making capability. Therefore, learning a single universal model for device-specific tasks is neither reasonable nor realistic. To adequately address these problems, it is necessary to allow each device to learn and use a *local, personalized* model [6]. In some other applications, it is convenient for groups of clients, called clusters, to share a single model, i.e., a personalized model for a cluster instead of an individual client. Further, it is also expected that the underlying device-specific models will change over time in real-life applications. Close monitoring of these changes is necessary to make appropriate personalized decisions. Therefore, FL techniques intended for real-life applications must be capable of adjusting their underlying models as new data becomes available.



$\mathbf{x}_{k,n}$: streaming datat at client k, $\mathbf{w}_{k,n}$: kth client local model
$\mathbf{w}^*_{k,n}$: enhanced personalized model

Fig. 2. Personalized online federated learning.

The exciting aspect of these device-specific tasks is that they differ but may be related, i.e., some similarities may exist between the tasks performed by individual clients or clusters. It is often necessary to take advantage of these similarities when a limited amount of data is available at a single client or cluster to build satisfactory personalized models. Promoting similarities in device-specific tasks during the learning process enables every client or cluster to build their best-customized models tailored to their own/local needs.

Personalized Online-FL meets the requirements of all the abovementioned cases better than traditional FL. Fig. 2 illustrates the workflow of personalized Online-FL. In personalized Online-FL, clients perform local learning on their own streaming data $\mathbf{x}_{k,n}$ to build personalized models $\mathbf{w}_{k,n}$ and communicate them to the server. Since the data is only available at irregular intervals to clients, local learning also happens irregularly. The server performs cluster analysis on the received models to group the clients into clusters, wherein the clients that belong to the same cluster strive to learn the same model. By using the cluster information, the server produces improved cluster-specific models by performing intra-cluster cooperation, i.e., averaging the local models that belong to the same cluster. The server finally generates enhanced personalized models $\mathbf{w}^\star_{k,n}$ by performing inter-cluster cooperation, i.e., promoting similarities between cluster-specific models. In the abovementioned personalized Online-FL workflow, intra-cluster cooperation will not be present if the number of clusters equals the number of clients. In contrast, if all the clients are part of the same cluster, there will be no inter-cluster cooperation, and the personalized Online-FL will be reduced to the traditional Online-FL.

We examined personalized Online-FL by applying it to a linear regression problem, using a synthetic dataset for training and the mean squared error (MSE) as a performance measure. We considered a scenario in which 100 clients, randomly grouped into 3 clusters, are connected to a server. Each cluster of clients aimed to learn a cluster-specific model of length 200. The server randomly selects 10% of the clients to participate in every global iteration. Every client performs local learning on a heterogeneous data stream using the stochastic gradient descent rule. Both traditional and personalized Online-FL learning methods were studied in this simulation example. Their learning performances are compared as shown in Fig. 3, in which each curve represents the MSE averaged over all clients for each method. The results demonstrate the superiority of personalized Online-FL over traditional Online-FL.

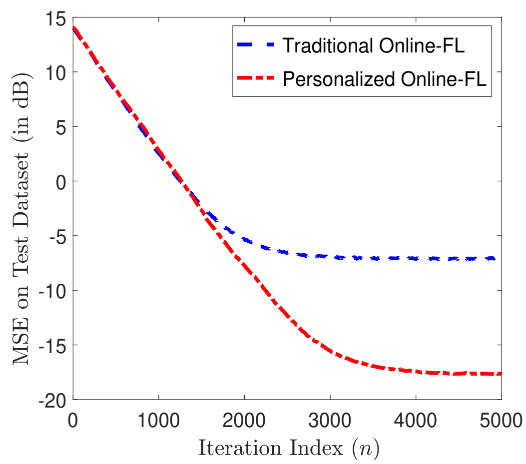


Fig. 3. Performance of personalized Online-FL vs. traditional Online-FL.

III. Applications

This section briefly describes six real-life applications that personalized Online-FL can efficiently serve. Of course, many other applications could benefit from personalized Online-FL, but the space constraint limits us to presenting these only.

A. Personalized Healthcare

IoT allows the healthcare system to expand its reach beyond the traditional clinical setting. Various wearable biomedical IoT devices, such as fitness trackers, blood glucose and diabetes monitors, blood pressure meters, cardiac rhythm monitors, and many more, are used to monitor patients' health status in real-time. These IoT devices generate a stream of time series measurements, possibly at irregular intervals [7]. It is well known that health status and disease progression differ significantly from patient to patient and are highly affected by the weather, wealth, pollution, and other socioeconomic factors. Personalized models can capture these patient-specific dynamics more precisely. Unfortunately, the amount of data obtained from just one of these IoT edge devices is often insufficient to build reliable models to adequately describe patient-specific behavior. Furthermore, acquiring healthcare data from biomedical devices owned by other users or hospitals can be challenging because laws restrict access and disclosure to protect patient privacy. Personalized Online-FL offers greater flexibility over traditional FL in this scenario, i.e., it enables edge devices to learn intelligent, personalized healthcare models from heterogeneous streams of healthcare data collected by biomedical IoT edge devices. Hence, each participating edge device could benefit from data collected by other devices beyond its own but without disclosing private information. A significant benefit of personalized Online-FL is that it learns personalized healthcare models to identify and diagnose each patient's disease accurately (and more effectively) by exploiting the similarities between the patient's health behaviors, regardless of whether they belong to the same or different geographical regions. These benefits of personalized Online-FL will tremendously impact healthcare in the future.

B. Networked Vehicles

In the automotive industry, autonomous driving technology is growing in popularity; its sound performance is crucial to the safety of the public as well as the economic prosperity of the industry. Autonomous vehicles need to be always aware of their surroundings. They need to learn a real-time personalized model evolving in accordance with their highly dynamic environment [8], thereby making personalized decisions on multiple tasks, such as steering prediction, lane detection, vehicle detection, and pedestrian detection. However, the decision-making ability of autonomous vehicles is limited when these models are learned just on their own data collected continually from a set of cameras and sensors. Those vehicles can make better decisions when they exchange information about their surroundings with neighboring cars, smart city devices equipped with cameras, and hand-held devices of pedestrians (e.g., smartphones and watches). Through this information exchange, each autonomous vehicle can obtain a full 3D model of its surroundings even in crowded environments, significantly improving effectiveness of decision-making. This data exchange may, however, raise privacy concerns. Personalized Online-FL is a solution tailored for this type of problems. It enables vehicles to learn vehicle-specific time-varying environment models from heterogeneous data streams and leverage the overlap between vehicles' local environment models while preserving users' privacy.

C. Smart Homes

IoT-enabled smart home environments are gaining popularity as they improve comfort and ease of use in everyday life. Household IoT devices, including smart TVs, cameras, speakers and microphones, air conditioners, lights, doors, windows, and many others, share environments and communicate via Wi-Fi. These smart devices are designed to monitor their local environment continually and perform only a specific task in response to any local change or local demand. However, these devices are vulnerable to cyberattacks which might cause physical harm. Context-aware control policies that allow or block a particular IoT access are recommended to prevent privacy leakage and physical hazards [10]. User-specific contextual policies, however, cannot be covered by manually generated policies. Instead, we may develop a user-specific contextual access control model using machine learning (ML) algorithms. Some smart homes may have limited data, which may hamper the accuracy of contextual access control models. Given each device's limited power and privacy reasons, it is preferable only to store and update the contextual access control models locally. Enhanced user-specific contextual access control models can be learned through personalized Online-FL without compromising the privacy of smart homes participating in the learning. Personalized Online-FL can also be used to learn user-specific speaker recognition models that provide access to control smart home appliances.

D. Precision Agriculture

Unlike autonomous vehicles, where legal restrictions limit innovation potential, autonomous mobile devices are flourishing in other fields. For instance, autonomous mobile devices for precision agriculture are receiving growing interest as they provide an alternative to polluting herbicides. Laser-equipped small, autonomous, rover-like devices are being used to dispose of weeds in fields. However, such devices have limited autonomy because of their size and battery capacity; therefore, they must be used in groups to handle large fields. In such cases, these devices need to recognize undesirable weeds and collaboratively learn the field's shape and device-specific path in real-time without impacting the performance of other devices. To this end, a lab-made base model of undesirable weeds can be uploaded to each device. Then, the autonomous mobile devices launched in a field would learn the personalized models that describe the shape of the field, their path, and the encountered weeds in real-time through observations and interactions among them. These personalized models would allow each device to know the position of all devices and the portion of the field already covered and collaboratively plan the best route by exchanging directions. A solution tailored for this type of applications is personalized Online-FL.

E. Personalized Digital Immersion

The ability to keep users immersed for a more extended period is critical to the success of any digital game. A digital game can become a popular favorite among a wider audience if its core elements are tailored to individual users' interests [9]. Therefore, personalization is an essential aspect of gaming. Personalization can take different forms. Some users are very particular about character aesthetics in games, such as vintage clothing, hairstyles, etc. Others may prefer playing against formidable opponents who challenge their abilities, enhancing their overall gaming experience and adding more enjoyment and excitement. Moreover, a given user's preferences may change with time. Those aforementioned characteristics of video games thus impose a need for gaming industries to continually monitor their users' behavior, abilities, and gaming styles to adapt to their preferences. Personalized Online-FL will serve the gaming industry's needs better than traditional FL, allowing it to learn time-varying user-specific gaming preferences from heterogeneous streams of users' gaming activities without revealing personal information. In addition to gaming, personalized Online-FL can also enhance the personalized experience in storytelling.

F. Wireless Systems

The use of reconfigurable intelligent surfaces (RISs) (also called intelligent reflecting surfaces) is increasingly becoming popular as they enhance the performance of wireless systems. RISs reflect signals from a base station (BS) toward users by controlling the coefficients of the RIS elements. In this way, RISs boost the received signal energy for remote users and extend the coverage area of the BS [11]. The RIS-assisted wireless systems utilize channel estimates to design the phase shifts of the reflecting beamformer elements. By doing so, the learning mechanism can select the best possible channels for the end user to communicate with the base station at any time. The performance of RIS-assisted wireless systems strongly relies on the accuracy of the instantaneous channel state information. However, RIS-assisted wireless systems involve signal reception through multiple channels that vary over time, which makes the channel estimation task more challenging. Although the edge devices communicating with a BS estimate their channels, sharing the similarity among channels will improve the estimation accuracy. With personalized Online-FL, it is possible to fuse the network information securely and intelligently while providing each user with a specific time-varying channel model corresponding to its location and the number of required channels.

IV. Challenges and Future Directions

This section discusses six key challenges that need to be addressed to implement successfully personalized Online-FL in the aforementioned real-life applications. Additionally, we present potential future directions for addressing these challenges.

A. Multi-Server Architectures

Although FL offers flexibility over centralized learning, traditional and personalized Online- FL frameworks involve many clients learning from their data and communicating their updated local models to a server. The server aggregates the local models to produce a global shared model that will be transferred back to clients. However, reliance on a single server is a significant limitation. If a given system is poorly scaled or unexpectedly stimulated, the large number of clients and the high dimension of the models to be shared would strain the communication channels and can lead to blockages. Furthermore, the complex and refined aggregation mechanisms handling many local models at the server might create computational bottlenecks, thus limiting the learning speed. In some specific applications, the geographical scarcity of clients does not allow a single server to possess enough data to aggregate models adequately. This device dispersion is especially likely in personalized learning, where several models must be learned. A natural solution studied in relatively few works is client-edge-cloud hierarchical FL. Client-edge-cloud hierarchical FL maintains multiple edge servers linked to their respective clients. After performing partial model aggregation, these edge servers communicate with a cloud server for final aggregation, which may suffer from the aforementioned problems and can become a single point of failure.

Recently proposed graph federated learning (GFL) comprises several interconnected servers, each associated with its own set of clients [12]. In GFL, learning is done in

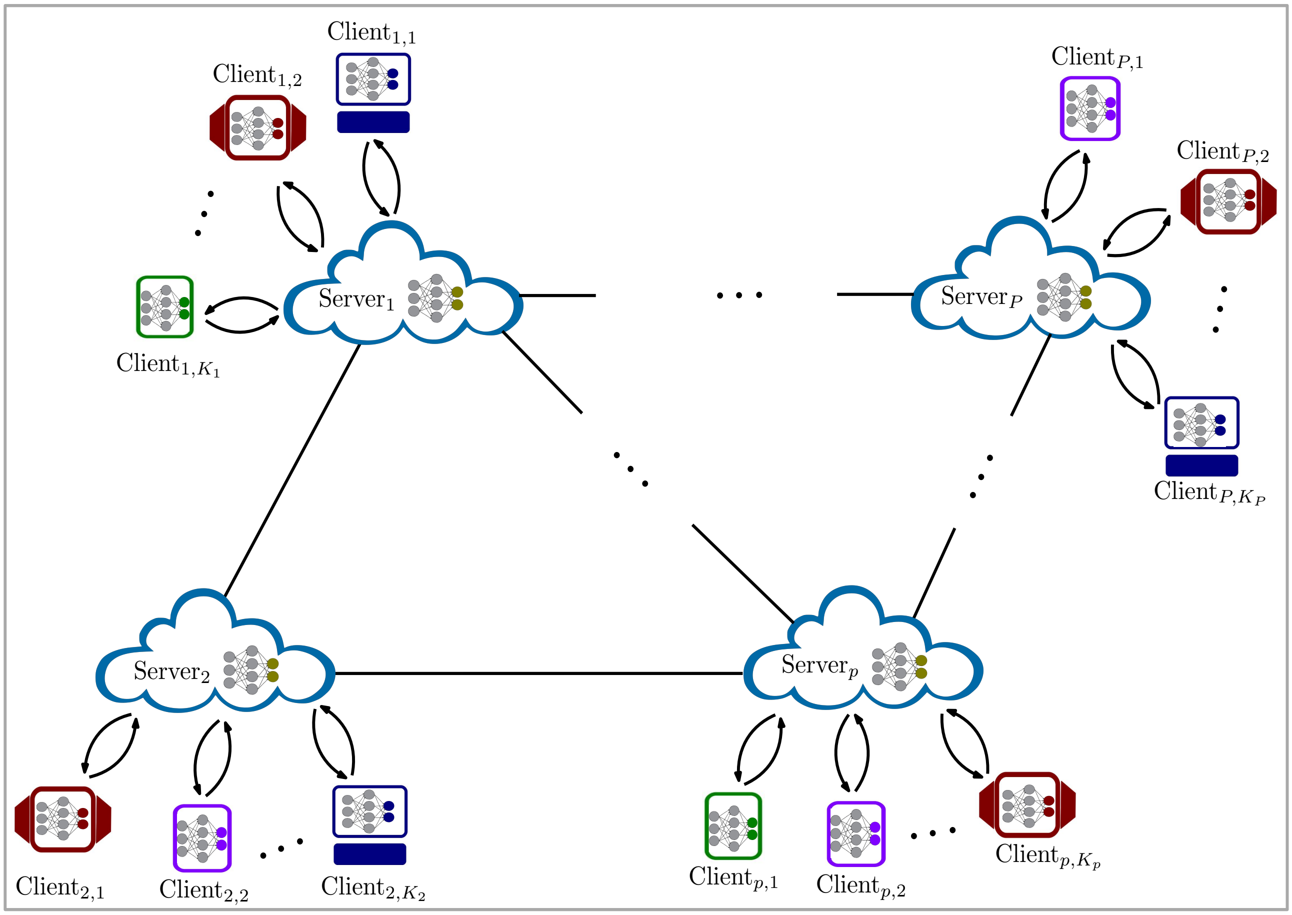


Fig. 4. Personalized Online-FL using graph federated architecture with P number of servers. Each server p is connected to its K_p number of clients.

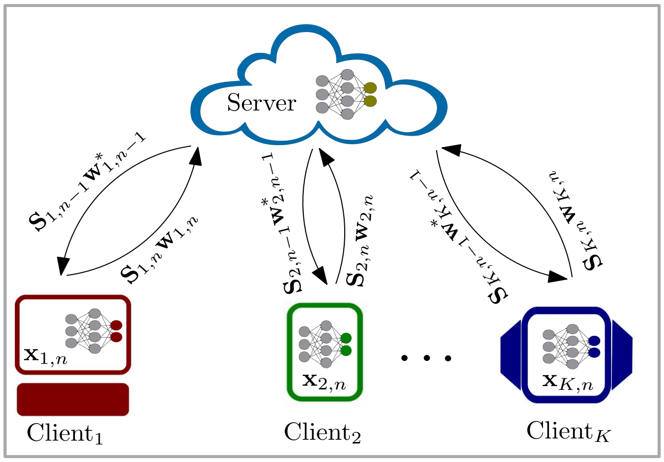
two stages. Firstly, each server performs personalized online learning with its respective clients, referred to as intra-server learning. Thereafter, the servers, amongst themselves, perform distributed personalized learning, referred to as intra-server learning. Unlike client-edge-cloud hierarchical FL, the servers in GFL perform learning in a distributed fashion. Hence GFL is suitable for practical applications. Because of these reasons, a GFL architecture is preferred over single-server architecture for an efficient, personalized Online-FL implementation. Fig. 4 illustrates the personalized Online-FL using graph federated architecture. During intra- and inter-server learning, the servers exploit the similarities among device-specific models via appropriate regularizers.

A graph federated architecture requires the selection of multiple computing servers or powerful clients (fog clients) as aggregation centers. Furthermore, each server receiving information from its clients and neighboring servers calls for a two-step or composed aggregation mechanism that accounts for the significance of the other server models. In unreliable systems, server-to-server communication links might need additional care or rely on clients to reach one another. Finally, careful clustering of the clients is necessary to ensure that a single client's model is not considered in many servers. These challenges must be studied when multi-server-based FL architectures are adopted for personalized Online-FL.

B. Communication-Efficiency

Communicating high-dimensional models back and forth to the server in FL requires more energy and bandwidth. IoT devices and sensors are typically resource-constrained, and have limited battery capacity. Consequently, communicating high-dimensional models in a short period might be challenging for them. In dense environments, competition among many devices for bandwidth can lead to communication bottlenecks, and strained communication channels may lead to malfunctions, such as blockages. Since the data arrive progressively in Online-FL, the models need to be updated and communicated whenever new data becomes available to the clients; this exacerbates the above limitations. Additionally, personalized Online-FL based on multiple servers involves two stages of model communication. In light of these issues, there is a great need for communication-efficient FL schemes.

There has been an increasing interest in communication-efficient FL, and efforts have been made to reduce the communication requirements associated with traditional FL. The two most prevalent approaches to reducing communication overhead are scheduling and compression. In scheduling, only a subset of the clients will be selected to participate in the learning at each global iteration, ensuring reduced strain on the communication channels at the cost of data variety. In compressed update methods, the local models are sparsified



$\mathbf{x}_{k,n}$: streaming data at client k, $\mathbf{w}_{k,n}$: kth client local model, $\mathbf{w}^*_{k,n}$: enhanced personalized model, $\mathbf{S}_{k,n}$: selection matrix for kth client

Fig. 5. Personalized Online-FL using partial-sharing-based communication.

and projected into a lower dimension to reduce the length of the messages. This compression reduces the communication overhead without sacrificing data variety; however, there is an innate accuracy cost associated with the sparsification process, and the simultaneous unpacking of the data at the server might lead to computation bottlenecks. Another method explored recently in FL is partial-sharing-based communication [13], in which clients exchange only a tiny portion of their local model parameters at each global iteration as shown in Fig. 5. Time-evolving diagonal selection matrices $\mathbf{S}_{k,n}$, containing only ones and zeros on the principal diagonal, specify the model parameters that will be shared. These selection matrices keep track of shared model indices. As opposed to compressed communication, partial sharing-based communication does not incur any computational overhead. Compared to traditional FL methods, this method provides clients with greater control over local learning. Moreover, it is, to a certain extent, inherently resilient to so-called byzantine attacks, wherein trusted devices try to disrupt learning.

Communication-efficient mechanisms must be examined in relation to personalized Online-FL, where further optimization can be done in addition to the techniques listed above. Recalling that the client- or cluster-specific models are different but related, every client must identify the portions of the model that are client- or cluster-specific. It is, therefore, possible to reduce the number of communication resources dedicated to learning the common portion in personalized models over time while increasing the number of resources dedicated to learning the client-specific portion in personalized models. Furthermore, event-triggered mechanisms minimize excessive processing and communication overhead. By verifying the innovation of the newly available data, these mechanisms allow clients to selectively update local models and communicate those to the respective server only when it is beneficial. These event-triggered mechanisms provide significant benefits to clients with low computational resources.

C. Straggler Clients

It is common to come across straggling devices in most real-world applications relying on a network of distributed devices. For example, in networked vehicles, a vehicle may intentionally disconnect for various reasons or accidentally disconnect due to weak signaling while in a certain location. Stragglers represent devices that incorporate many limitations a device might have, such as low power availability, modest computational capability, imperfect communication channels, and susceptibility to failure. Due to these limitations, straggler devices impair learning, even more so in personalized Online-FL. As a result, different challenges will emerge in the context of personalized Online-FL. For instance, when a device working toward a personalized model becomes unresponsive for a particular period, one can not easily determine whether its model has strayed from the average or is outdated. Furthermore, given the limited size of client clusters, any misbehavior or lack thereof of a client can significantly impact the models of the other clients. For these reasons, practical straggler-related issues and model freshness should be addressed in personalized Online-FL. Various strategies have been developed to handle straggler devices in traditional FL, e.g., reactive and adaptive aggregation mechanisms, but very few have been extended to personalized FL. Hierarchical learning is a promising solution to improve the learning conditions in systems with stragglers. For instance, more robust clients with access to higher volumes of quality data participate in a partial model aggregation. They then share a reasonable server model with straggler clients to allow them to participate sporadically. This procedure can handle the poor computational capacity of stragglers. However, sometimes stragglers also have access to large volumes of quality data. Therefore, novel methods need to be developed to address this critical situation, e.g., by allowing straggler clients to sift through the data resource-efficiently while not losing important information whenever they contribute to global learning. Additionally, event-triggered learning mechanisms may benefit stragglers by reducing their computational burden.

D. Model-Poisoning Attacks

In many distributed learning applications, adversaries may try to disrupt the learning process by sharing random or purposefully harmful data with the participating devices, referred to as a model-poisoning attack [14]. In traditional FL and Online-FL, these attacks are often detected; consequently, those updates will be ignored by looking for outliers among the participating devices. Model-poisoning attacks may have additional harmful effects on personalized Online-FL. In personalized Online-FL, the devices are allowed to learn device-specific models, so outliers are an integral part of the process and cannot be removed when detected under the suspicion of being an adversary. A simple possible solution is allowing an adversary client to operate with a personalized model in the network, only concentrating on negating its impact on learning other client models. To this end, only a selected group of clients would be used to aggregate the base server model; the others would only benefit from this model and update their personalized models. Recently proposed sign stochastic gradient sharing- and partial-sharing-based communication are proven to be resilient against model-poising attacks. These techniques can be incorporated into the personalized Online-FL to address model-poising attacks. However, their ability to learn personalized models is yet to be studied. Furthermore, the resilience of these communication techniques has to be studied under stealth attacks.

E. Algorithm and Architecture Co-Design

Both in traditional FL and personalized Online-FL, we move computations to the network edge, i.e., the data processing takes place close to the point of its origin. The computation on the network edge brings opportunities for customization of algorithms and the device hardware to reduce latency and power demands. Models based on contemporary ML algorithms are typically large. Thus, FL's training phase involves many iterations before finalizing the model. Therefore, the local learning on the edge device demands higher energy and throughput. However, IoT edge devices are equipped with limited battery and computational power. These issues can be addressed with novel online ML architectures that handle heterogeneous data streams and have fewer neural network layers. However, ML algorithms for heterogeneous data streams are still in their infancy. A few recent works [15] have attempted this; nevertheless, much research needs to be done in this direction. When implementing these algorithms in hardware, an application-specific integrated circuit (ASIC) or field-programmable gate array (FPGA) needs to be considered to take advantage of recent developments in low-power, very large-scale integration (VLSI) architecture designs. These dedicated VLSI architectures can meet the higher throughput and lower latency requirements of real-time processing. Furthermore, the best trade-off between learning performance and implementation flexibility can be achieved by jointly customizing the algorithm and architecture rather than designing each separately.

F. Standardization

Finally, standardization must be addressed to make personalized Online-FL a reality. Many existing works assume precise settings regarding client distribution and predetermined clusters that often differ from one work to another. For instance, in some existing works, the cluster distribution of clients remains constant during the learning process. This assumption indeed results in mathematically tractable and computationally efficient solutions. However, it also closes the door to personalized Online-FL with adaptable clusters that could take advantage of the advancements in the field. Therefore, it is necessary to study and incorporate generic or adaptable cluster identification methods and intelligent aggregation methods into a standard framework to allow the smooth implementation of future works on solid foundations. Moreover, other features that deal with communication efficiency, privacy protection, asynchronous behavior of clients and communication channels, and adversary detection should not restrict the framework's compatibility. A combination of many of these improvements in a single framework is necessary as real-life applications often have several requirements of this type.

V. Concluding Remarks

Personalized Online-FL allows geographically dispersed edge devices to learn device-specific models from the event-triggered heterogeneous stream of measurements; it is, therefore, more suitable for IoT/CPS-based applications. Combining continuous learning and personalized models ensures that each device can access an up-to-date solution for its specific task. Thus, personalized Online-FL is an efficient alternative to traditional FL while retaining its collaborative and privacy-preserving nature. This article highlighted some of the many applications that could benefit from personalized Online-FL. Furthermore, we elaborated on the challenges that need to be overcome for personalized Online-FL to be widely utilized in practice. We also presented some potential solutions to address those challenges, intending the further research and development of personalized Online-FL.

References

[1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. and Stat.*, 2017, vol. 54, pp. 1273-1282.

[2] A. Mitra, H. Hassani and G. J. Pappas, "Online federated learning," in *Proc. IEEE Conf. Decision and Control*, 2021, pp. 4083-4090.

[3] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50-60, May 2020.

[4] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031-2063, 2020.

[5] T. Zhang, L. Gao, C. He, M. Zhang, B. Krishnamachari and A. S. Avestimehr, "Federated learning for the internet of things: Applications, challenges, and opportunities," *IEEE Internet of Things Mag.*, vol. 5, no. 1, pp. 24-29, Mar. 2022.

[6] A. Z. Tan, H. Yu, L. Cui and Q. Yang, "Towards personalized federated learning," *IEEE Trans. Neural Net. and Learning Syst.*, doi: 10.1109/TNNLS.2022.3160699.

[7] S. Boll and J. Meyer, "Health-X dataLOFT: A sovereign federated cloud for personalized health care services," *IEEE MultiMedia*, vol. 29, no. 1, pp. 136-140, 2022.

[8] B. Yang, X. Cao, K. Xiong, C. Yuen, Y. L. Guan, S Leng, L. Qian and Zhu Han, "Edge intelligence for autonomous driving in 6G wireless system: Design challenges and solutions," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 40-47, Apr. 2021.

[9] M. C. Angelides and H. Agius, *Handbook of Digital Games*, Wiley, 2014.

[10] T. Yu, T. Li, Y. Sun, S. Nanda, V. Smith, V. Sekar, and S. Seshan, "Learning context-aware policies from multiple smart homes via federated multi-task learning," in *Proc. IEEE/ACM Fifth Int. Conf. Internet-of-Things Design and Implementation*, 2020, pp. 104-115.

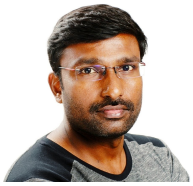
[11] S. Noh, J. Lee, G. Lee, K. Seo, Y. Sung and H. Yu, "Channel estimation techniques for RIS-assisted communication: Millimeter-wave and sub-THz systems," *IEEE Vehicular Technology Mag.*, vol. 17, no. 2, pp. 64-73, Jun. 2022.

[12] E. Rizk and A. H. Sayed, "A graph federated architecture with privacy preserving learning," in Proc. IEEE Int. Workshop on Signal Process. Advances in Wireless Commun., 2021, pp. 131-135.

[13] V. C. Gogineni, S. Werner, Y. Huang and A. Kuh, "Decentralized graph federated multitask learning for streaming data," in *Proc. Annual Conf. Info. Sci. and Syst.*, 2022, pp. 101-106.

[14] X. Ma, X. Sun, Y. Wu, Z. Liu, X. Chen and C. Dong, "Differentially private byzantine-robust federated learning," *IEEE Trans. on Parallel and Distributed Syst.*, vol. 33, no. 12, pp. 3690-3701, Dec. 2022.

[15] X. Zheng and V. Lau, "Federated online deep learning for CSIT and CSIR estimation of FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 70, pp. 2253-2266, Apr. 2022.



Vinay Chakravarthi Gogineni (Member, IEEE) received the Bachelor's degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Andhra Pradesh, India, in 2005, the Master's degree in communication engineering from VIT University, India, in 2008, and the Ph.D. degree in electronics and electrical communication engineering from Indian Institute of Technology Kharagpur, India in 2019. From 2008 to 2011, he was with a couple of MNCs in India. Currently, he is working as a postdoctoral research fellow at the department of electronic systems, NTNU-Norway. His research interests include statistical signal processing, distributed and federated learning, and geometric deep learning. He was a recipient of the ERCIM Alain Bensoussan Fellowship in 2019 and the Best Paper Award at APSIPA ASC-2021, Tokyo, Japan.

Stefan Werner (Senior Member, IEEE) received the M.Sc. Degree in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden, in 1998, and the D.Sc. degree (Hons.) in electrical engineering from the Signal Processing Laboratory, Helsinki University of Technology, Espoo, Finland, in 2002. He is currently a Professor at the Department of Electronic Systems, Norwegian University of Science and Technology (NTNU), Director of IoT@NTNU, and Adjunct Professor with Aalto University in Finland. He was a visiting Melchor Professor with the University of Notre Dame during the summer of 2019 and an Adjunct Senior Research Fellow with the Institute for Telecommunications Research, University of South Australia, from 2014 to 2020. He held an Academy Research Fellowship, funded by the Academy of Finland, from 2009 to 2014. His research interests include adaptive and statistical signal processing, wireless communications, and security and privacy in cyber-physical systems. He is a member of the editorial boards for the EURASIP Journal of Signal Processing and the IEEE Transactions on Signal and Information Processing over Networks.

Francois Gauthier (Member, IEEE) received both the B.Sc. and M.Sc. degree in mathematics and computer science from École Nationale Supérieure d'Informatiques et de Mathématiques Appliquées de Grenoble. He is pursuing the Ph.D degree at the department of Electronic Systems in the Norwegian University of Science and Technology (NTNU), Trondheim, Norway. His research interests are federated learning, differential privacy, asynchronous behavior, and communication efficiency.

Yih-Fang Huang (Life Fellow, IEEE) is Professor of Electrical Engineering and Senior Associate Dean for Education and Undergraduate Programs in the College of Engineering. He received his B.S.E.E. degree from National Taiwan University, M.S.E.E. degree from University of Notre Dame, M.A. and Ph.D. degrees from Princeton University. He served as chair of Notre Dame's Electrical Engineering department from 1998 to 2006. His research lies in the area of statistical and adaptive signal processing and employs principles in mathematical statistics to solve signal detection and estimation problems that arise in various applications, including wireless communications, distributed sensor networks, smart electric power grid, etc.

Dr. Huang received the Golden Jubilee Medal of the IEEE Circuits and Systems Society in 1999. He also served as Vice President in 1997-98 and was a Distinguished Lecturer for the same society in 2000-2001. He served as the lead Guest Editor for a Special Issue on Signal Processing in Smart Electric Power Grid of the IEEE Journal of Selected Topics in Signal Processing, December 2014. At the University of Notre Dame, he received Presidential Award in 2003, the Electrical Engineering department's Outstanding Teacher Award in 1994 and in 2011, the Rev. Edmund P. Joyce, CSC Award for Excellence in Undergraduate Teaching in 2011, and the Engineering College's Outstanding Teacher of the Year Award in 2013.

In Spring 1993, Dr. Huang received the Toshiba Fellowship and was Toshiba Visiting Professor at Waseda University, Tokyo, Japan. From April to July 2007, he was a visiting professor at the Munich University of Technology, Germany. In Fall, 2007, Dr. Huang was awarded the Fulbright-Nokia scholarship for lectures/research at Helsinki University of Technology in Finland. He was appointed Honorary Professor in the College of Electrical Engineering and Computer Science at National Chiao-Tung University, Hsinchu, Taiwan, in 2014. Dr. Huang is a Life Fellow of the IEEE.

Anthony Kuh (Fellow, IEEE) received his B.S. in Electrical Engineering and Computer Science at the University of California, Berkeley in 1979, an M.S. in Electrical Engineering from Stanford University in 1980, and a Ph.D. in Electrical Engineering from Princeton University in 1987. He previously worked at AT&T Bell Laboratories and has been on the faculty in the Department of Electrical and Computer Engineering at the University of Hawai'i since 1986. He is currently a Professor and previously served as Department Chair. His research is in the area of neural networks and machine learning, adaptive signal processing, sensor networks, and renewable energy and smart grid applications. He won a National Science Foundation (NSF) Presidential Young Investigator Award and is an IEEE Fellow. From 2017-2021, he served as program director for NSF in the Electrical, Communications, and Cyber Systems (ECCS) division working in the Energy, Power, Control, and Network (EPCN) group. At NSF he also assisted in initiatives including Harnessing the Data Revolution (HDR), the Mathematics of Deep Learning (MoDL), the AI Institutes, Cyber Physical Systems (CPS), and Smart and Connected Communities. He previously served for the IEEE Signal Processing Society on the Board of Governors as a Regional Director-at-Large Regions1-6, as a senior editor for the IEEE Journal of Selected Topics in Signal Processing, and as a member of the Awards Board. He currently is the President of the Asia Pacific Signal and Information Processing Association.