A Survey on Participant Selection for Federated Learning in Mobile Networks

Behnaz Soltani[1], Venus Haghighi[1], Adnan Mahmood[1], Quan Z. Sheng[1], Lina Yao[2]
[1]Macquarie University, Sydney, Australia
[2]University of New South Wales, Sydney, Australia

ABSTRACT

Federated Learning (FL) is an efficient distributed machine learning paradigm that employs private datasets in a privacy-preserving manner. The main challenges of FL are that end devices usually possess various computation and communication capabilities and their training data are not independent and identically distributed (non-IID). Due to limited communication bandwidth and unstable availability of such devices in a mobile network, only a fraction of end devices (also referred to as the *participants* or *clients* in a FL process) can be selected in each round. Hence, it is of paramount importance to utilize an efficient participant selection scheme to maximize the performance of FL including final model accuracy and training time. In this paper, we provide a review of participant selection techniques for FL. First, we introduce FL and highlight the main challenges during participant selection. Then, we review the existing studies and categorize them based on their solutions. Finally, we provide some future directions on participant selection for FL based on our analysis of the state-of-the-art in this topic area.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; **Distributed computing methodologies**.

KEYWORDS

Federated learning, machine learning, participant selection.

ACM Reference Format:
Behnaz Soltani[1], Venus Haghighi[1], Adnan Mahmood[1], Quan Z. Sheng[1], Lina Yao[2]. 2022. A Survey on Participant Selection for Federated Learning in Mobile Networks. In *17th ACM Workshop on Mobility in the Evolving Internet Architecture (MobiArch'22), October 21, 2022, Sydney, NSW, Australia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3556548.3559633

MobiArch'22, October 21, 2022, Sydney, NSW, Australia

ACM ISBN 978-1-4503-9518-2/22/10…$15.00
https://doi.org/10.1145/3556548.3559633

1 INTRODUCTION

Mobile devices such as vehicles and smart phones are constantly generating a massive amount of data, which could be utilized for machine learning in a bid to achieve smart mobile applications. However, transmitting private data to a centralized or an edge server for training may lead to privacy issue and could cause long communication latency and large resource cost [1].

A decentralized machine learning approach called Federated Learning (FL) has been proposed by Google that enables cooperative learning on devices without sharing the local data [2]. Clients train the model on-device in a privacy-preserving manner using their local datasets and transfer the local model parameters to the FL server for aggregation. As a result, FL enables user privacy preservation, low communication costs, and transmission latency reduction owing to transmitting only model parameters to the server for aggregation. Furthermore, in time-critical systems such as autonomous vehicles, making real time decisions locally at end devices significantly decreases response time.

Employing various datasets of different clients using FL leads to model accuracy improvement. Participant selection is an emerging challenge in the management of FL that possesses a profound impact on the efficiency of the model training, especially in the scenarios with a huge number of participants and limited wireless channels. In FL, due to dynamic environments and the limited network bandwidth, only a fraction of clients can participate for training in each round. Hence, proper selection is of paramount importance in FL to achieve the desired accuracy with fast convergence. In FedAvg [2] algorithm introduced by Google, participants are determined uniformly at random in each round. However, due to the various computation and communication resources and heterogeneous datasets, devices may have different contributions on the training performance. To evaluate the performance of FL, *training time* and *final model accuracy* are the most important factors. Mostly, there are trade-offs between the model accuracy and convergence time in FL. For example, participants with high quality data may

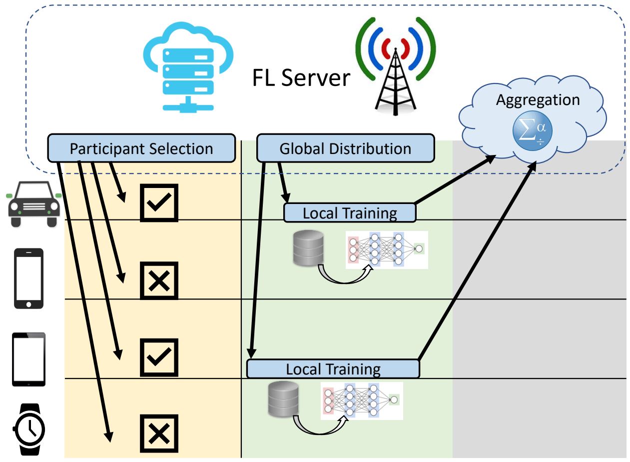

Figure 1: An Overview of Federated Learning Process.

have poor network connection or computation capacities, while those with low quality data may possess rich resources to perform training process.

To the best of our knowledge, there are no existing works that conduct a literature review of the participant selection in FL. The main contributions of this study are as follows:

- We discuss the important challenges pertinent to participant selection process in FL.
- We review the state-of-the-art on participant selection by categorizing them vis-à-vis different approaches.
- We identify some open research directions of participant selection in mobile networks.

2 FEDERATED LEARNING

Federation learning (FL), in essence, encompasses a number of steps which are delineated as follows:

(1) *Initialization* – The aggregation server determines the FL task, generates a randomly or pretrained global model, and adjusts learning parameters (e.g., the number of rounds and the learning rates).
(2) *Participant Selection* – The FL server selects n number of clients amongst the volunteers to participate in the training process.
(3) *Global Distribution* – The server disseminates the global model parameters to the selected participants to train and update the shared model.
(4) *Local Training* – Selected participants train the shared model using their local data samples and upload local model parameters to the server.
(5) *Aggregation* – The server aggregates local models and computes a new global model.

Steps 2, 3, and 4 repeat until the model converges or reaches a desirable accuracy and are portrayed in Figure 1.

The most popular FL algorithm is FedAvg [2] that randomly samples a subset of clients. However, due to different computation and communication capabilities and various data samples, randomly selected participants could degrade the performance of FL process.

2.1 Challenges of Participant Selection

There are several significant challenges in participant selection approaches that affect the performance of FL process:

- **Device Heterogeneity:** Devices have different computation, communication, and storage capabilities that may negatively affect the performance of training process [3]. In synchronous FL with heterogeneous resources, per-round training time is determined by the slowest clients (i.e., *stragglers*) since all the participants are required to wait for the slowest clients to complete their training tasks. On the other hand, end devices with limited resources may increase dropouts during training process and affect the accuracy of the model. Therefore, selecting clients without considering their resource capabilities may lead to longer training time and degrade the final model accuracy.
- **Data Heterogeneity**: Since data is generated independently based on clients' behaviours, data distribution on end devices are usually non-Independent and Identically Distributed (non-IID). Hence, local datasets may not be representative of the population distribution, which can lead to the biased model update and degrade the model accuracy [2, 4–6].
- **Dynamicity:** Clients might be unavailable due to the high-mobility environment, poor network condition, and energy constraints [7–9]. Furthermore, due to the channel fading in wireless networks, a fraction of local model updates may be lost. Therefore, a dynamic environment with high mobility devices has a major impact on the performance of FL process.
- **Trustworthiness:** Since the FL server has no knowledge about the local training process, malicious devices may launch attacks and manipulate the result of the training task. Therefore, it is of utmost importance to identify malicious clients and eliminate them from the learning process.
- **Fairness:** Devices with poor capabilities are less likely to be selected to participate in the training process, which leads to selection bias and degrade the model accuracy [10]. Fairness enables clients with diverse datasets to participate in the FL process and improve the model accuracy and convergence speed. Therefore, in order to minimize the model bias and generalize the global model, all the end devices should have a chance to participate in the FL process.

3 PARTICIPANT SELECTION IN FL

Over the past few years, researchers in both academia and industry have proposed a number of research studies pertinent to participant selection in the FL process. The same have been classified into eight appropriate categories and are discussed as follows:

Threshold-based Selection. Participants may be allowed to complete the local training within a specific deadline. FedCS [11] is one of the first studies on client selection in FL that manages participant selection process based on the resource capabilities. The proposed method aims to select as many clients as possible that can complete the training steps within a specific deadline to reach the desired performance [2]. First, the random clients are asked to inform their resource information, i.e., computation and computation capacities, wireless channel states, and the size of their data. Afterward, the server computes the required time to complete the training task for each client using aforementioned information. However, due to dynamically changing resource and network conditions, the static time threshold cannot guarantee the efficiency of this approach. Thus, an adaptive deadline determination algorithm for mobile devices is proposed in [12], wherein clients are adaptively determined at each round instead of using a fixed deadline. This approach could decrease the convergence time by up to 50% in contrast to the conventional fixed threshold schemes.

A multicriteria-based client selection strategy is proposed in [13] in which time, CPU, memory, and energy for all the end devices are considered to predict their respective capability to perform a training task. The authors employ stratified-based sampling to establish a homogeneous group of clients based on the time zones. Subsequently, considering the resource utilization and the resource capabilities to complete the FL task within a specific time frame, the authors formulate a bilevel optimization problem to maximize the number of selected clients. To predict the resource utilization, linear regression according to the history of the past rounds is employed. Furthermore, due to imbalanced class distribution, the authors prioritize participants with the highest event rate (i.e., the ratio of abnormal data samples pertaining to minority class to the total data samples)

Reputation-based Selection. Unreliable clients may (a) perform undesirable behaviours intentionally, e.g., data poisoning, model poisoning, and sybil attacks, or (b) unintentionally due to high mobility environments and unstable network connections. Therefore, it is of paramount importance to design an efficient participant selection approach for reliable model training. Reputation is a metric to evaluate trustworthiness of entities based on their past interactions [14–16]. Recently, a couple of studies employed reputation as a metric to select reliable devices for participating in FL process. A reputation-based client selection scheme is proposed in [8, 9] that employs multiweight subjective logic by taking into account both the direct interactions and recommendations from other task publishers. The proposed approach considers three weight attributes: interaction frequency, interaction timeliness, and interaction effects. The reputation is stored in a blockchain to achieve secure reputation management.

In [17], a reputation-based regional FL framework is introduced for intelligent transportation systems in which vehicles are divided into multiple regions each of which possesses its own learning model. Roadside Units (RSUs) are selected as the leaders and perform aggregation in each region. Owing to the dynamic nature of transportation systems, devices may move away from their regions and join a new region. Each leader computes the reputation of vehicles based on a) honesty degree, b) accuracy contribution, and c) interaction timeliness, and selects vehicles with high reputation to join FL process. In [18], the authors introduce a reputation model based on beta distribution in order to measure the trustworthiness of the end devices. The trust value is evaluated using the contribution of participants to the global model. They propose a reputation-based scheduling scheme that jointly considers trustworthiness and fairness based on the reputation value and successful transmission rate.

Probability Allocation-based Selection. Different end devices can have various probabilities of being selected for a training task. The authors in [3] design a client sampling strategy to reduce total training time. They take both system and data heterogeneity into consideration since clients with high-quality data may have poor communication resources, whereas, those with high communication capabilities may have low-quality data. Their approach provides a new convergence upper bound for arbitrary client selection probabilities and generates a non-convex training time minimization problem. Their strategy reduces the convergence time for achieving the same target loss compared to the baselines.

In [10], a stochastic client selection algorithm under a volatile context is investigated in which the selected clients might not be capable of returning their local models for aggregation due to different reasons, including but not limited to, limited computing resources, network failure, and user unwillingness. However, selecting clients with the lowest failure probability may violate selection fairness owing to the selection of a particular group of clients repeatedly and hurt the model accuracy. Therefore, the authors study the trade-off between selecting participants with the lowest failure likelihood and selection fairness, and formulate the client selection problem by taking both factors into account. Since the problem cannot be solved offline due to unknown status of the participants, an adversary bandit-based online scheduling solution is employed. The proposed method is able to accelerate convergence while maintaining the model accuracy level.

The authors in [19] derive a novel convergence upper bound for non-convex loss functions using FL with arbitrary device selection probabilities. They design a stochastic optimization problem to minimize a weighted sum of the convergence bound and communication time. The proposed selection method solves the problem using the Lyapunov drift-plus-penalty framework based on current channel conditions without the knowledge of channel statistics. In [20], an aggregation algorithm is designed to determine the optimal subset of local model updates by excluding adverse local updates. Moreover, a Probabilistic Node Selection framework (FedPNS) is proposed that dynamically adjusts the selection probability for devices based on their contribution to the model pertaining to the data distribution that is determined using the output of the aggregation algorithm.

Reinforcement Learning-based Selection. In reinforcement learning, an agent learns to attain an objective in a complex and uncertain environment and maximizes its rewards. A framework based on multi-armed bandit for online client selection is introduced in [21] to minimize the training latency including both local computation time and data transmission time in two scenarios: 1) an ideal scenario in which clients possess balanced and IID datasets and are always available, and 2) a non-ideal scenario in which clients may be unavailable and the distribution of datasets is non-IID. In the non-IID scenario, both the fairness and availability constraint is addressed.

In [22], the authors define an offline client selection problem with long-term fairness constraints. The constraint adjusts the minimum average selection rate of every participant to maintain fairness for the system. Due to the indeterminate availability of clients until the start of a round and also time-coupling fairness constraint, the authors utilize the Lyapunov optimization framework to transfer the offline problem into an online optimization problem, where the participation rate of clients is evaluated using dynamic queues. In the proposed method, the model exchange time (i.e., the time spent between global model distribution and uploading all the local models) of each participant before each communication round is estimated using Contextual Combinatorial Multi Arm Bandit (C^2MAB) model. The authors aim to minimize the long-term model exchange time.

An experience-driven controlled framework called FAVOR is proposed in [5] that aims to determine the best fraction of devices in each round to minimize the number of rounds and tackle the data non-IID distribution. The authors formulate client selection for FL process as a deep reinforcement learning problem. To select devices in each communication round, a double deep Q-learning mechanism is proposed in order to improve global model accuracy and reduce the number of communication rounds.

Group-based Selection Clients are divided into several groups and then those within the same group are selected for each communication round. In [23], a grouping based participant selection mechanism is introduced in which participants are split into various groups based on group earth mover's distance (GEMD) to balance the label distribution of the clients. This new metric evaluates similarity between global distribution and local data distributions. A smaller GEMD means that the training data of the selected clients are closer to IID distribution. Therefore, selecting a group of clients with the smallest GEMD can improve the performance of FL.

Weight Divergence-based Selection. Weight divergence in training can be an indicator to identify the distribution of local datasets. The authors in [24] introduce a client selection utility that tries to deal with the trade-off between accuracy and execution time in each round. The change of weight between two adjacent rounds is defined as a utility for fast convergence. In addition, since clients with large data volume may negatively affect the training time, the ratio of the local data size to the total size is also added as a coefficient to the client's utility in order that if local data constitutes a significant amount of total data, utility reduces. Since it is not always necessary to select participants in every round, this study also designs a feedback control component that dynamically adjusts the frequency of client selection.

The authors in [6] design a participant selection algorithm to tackle the accuracy reduction owing to non-IID datasets. They identify the degree of non-IID data using the weight divergence. The weight divergence is evaluated between the client model and auxiliary model which is trained using public or purchased datasets at the aggregation server. In this approach, participants that have lower non-IID degree should be selected with higher frequency for training.

Offloading-Aware Selection. Device-to-device offloading of data processing from resource-limited devices to resource-rich devices can enable local training with more diversified data distribution. Participant sampling with data offloading are combined in [25] to maximize training accuracy. Devices with large contributions to model training are chosen for training and other devices may send their data to the selected participants based on their data similarity. Two data samples are considered similar if they have identical labels and their feature vectors have limited difference. Data offloading is only performed between trusted and single-hop neighbors. To determine participants, the proposed method uses graph convolutional networks (GCNs) to learn the relationship between FL accuracy, network attributes, and offloading topology, which maximizes training accuracy.

Table 1: Comparison of the existing participant selection works.

No.	Methods	Device Heterogeneity	Data Heterogeneity	Fairness	Dynamicity	Trustworthiness	Objective
[11]	Threshold-based	✓	✗	✗	✗	✗	Maximizing the number of participants
[12]		✓	✗	✗	✓	✗	Maximizing the number of participants
[13]		✓	✓	✗	✗	✗	Maximizing the number of participants
[8, 9]	Reputation-based	✗	✗	✗	✗	✓	Slecting trusted participants to improve the reliablity of FL
[17]		✗	✗	✗	✓	✓	Reliable participant selection to improve accuracy of knowledge
[18]		✗	✗	✓	✓	✓	Improving the reliability and convergence performance
[3]	Probability allocation-based	✓	✓	✗	✗	✗	Minimizing total learning time
[10]		✓	✗	✓	✓	✗	Addressing the trade-off between the lowest failure probability and fairness
[19]		✓	✓	✗	✓	✗	Minimizing communication time for speeding up the convergence
[20]		✗	✓	✗	✗	✗	Excluding adverse local models from participating devices
[21]	Reinforcement learning-based	✓	✓	✓	✓	✗	Minimizing the total training time
[22]		✓	✗	✓	✓	✗	Minimizing exchange time with fairness guarantee
[5]		✗	✓	✗	✗	✗	Reducing the number of communication round under non-IID setting
[23]	Group-based	✗	✓	✗	✗	✗	Balancing the label distribution of participants
[24]	Weight Divergence-based	✗	✓	✗	✗	✗	Addressing trade-off between accuracy contribution and training time
[6]		✗	✓	✗	✗	✗	Selecting clients with lower non-IID degree of data
[25]	Offloading-Aware	✓	✓	✗	✗	✗	The optimal combination of client selection and data offloading configuration
[26]	Clustering-based	✗	✓	✗	✗	✗	Unbiased participant selection

Clustering-based Selection. Various clients can be selected with different data distributions using clustered sampling. In [26], an unbiased clustered sampling scheme for client selection is introduced that decreases weights' variance for the client aggregation and ensures that clients with unique distributions are more likely to be selected. The authors introduce two clustered sampling schemes: 1) clustered sampling based on sample size, and 2) clustered sampling based on similarity. They show that clustered sampling leads to better and faster convergence.

In Table 1, we summarize the existing studies and their approaches, and draw a comparison of them in terms of solving different challenges.

4 FUTURE RESEARCH DIRECTIONS

Participant selection in FL is still in its infancy and there are some open challenges that require to be addressed. We have identified the following main research directions:

Optimal and Adaptive Reputation Threshold. Existing reputation-based studies consider a pre-defined threshold in order to distinguish malicious devices from honest ones. If this threshold is adjusted too high, honest devices may lose the opportunity to participate in the FL process. On the other hand, if the aforementioned threshold is set too low, malicious devices may be selected for training and manipulate the model. It is, therefore, essential to optimize the threshold value, particularly, in safety-critical applications.

Hierarchical Aggregation. All the reviewed approaches use centralized aggregation, wherein clients are required to transmit their local model parameters to a single aggregator server. When the number of devices increase, limited network bandwidth becomes a scalability bottleneck in centralized FL. On the other hand, communication between end devices and a remote server might be intermittent or unavailable. In hierarchical FL, end devices are divided into a number of clusters and participants transmit their local models to the cluster head for intermediate model aggregation. All cluster heads communicate with the FL server for global aggregation [27]. This approach reduces the device dropout and improves scalability. To reduce communication cost and maximizing the training efficiency, it is necessary to develop an appropriate mechanism for determining the cluster structures.

Highly Dynamic Environment. In mobile networks such as Internet of Vehicles with high mobility nodes and a highly dynamic topology, intermittent connectivity of devices with the server along with a constant change in data transmission time may degrade the performance of FL. After client selection, some participants may get out of network coverage and significantly affect the system behaviour. It is important to handle high mobility environment and minimize the training performance degradation. Decentralized FL can be a proper solution for high mobility clients that may not have access to the aggregation server. In decentralized FL, clients can only share their local model parameters with nearby devices using device-to-device communication without relying on a single server [28]. Device selection algorithms in decentralized FL can have a significant impact on the training process. Social trust values can be integrated with other selection criteria to determine the best nearby clients.

Asynchronous Training. All existing works are based on synchronous training, wherein the aggregation server has to wait for receiving all the local models before starting aggregation process [29]. Therefore, *stragglers* prolong the training time since per-round training time is determined by the slowest clients. In asynchronous FL, the server is not affected by *stragglers* and can update the global model without waiting to collect all the local models. In addition, scalability can be improved using asynchronous FL [30]. Participant selection in asynchronous FL needs to be more explored in order to develop appropriate selection strategies to achieve desirable training performance.

5 CONCLUSION

Participant selection in federated learning (FL) has a significant impact on the final model accuracy and training time of FL. In this paper, we explore the main FL challenges of the participant selection process. We review the existing approaches and conduct a comparison of the existing studies in terms of solving different challenges. Finally, we identify some research directions and hope to stimulate further research in this important topic.

REFERENCES

[1] Joost Verbraeken, Matthijs Wolting, Jonathan Katzy, Jeroen Kloppenburg, Tim Verbelen, and Jan S Rellermeyer. A survey on distributed machine learning. *ACM Computing Surveys (CSUR)*, 53(2):1–33, 2020.

[2] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[3] Bing Luo, Wenli Xiao, Shiqiang Wang, Jianwei Huang, and Leandros Tassiulas. Tackling system and statistical heterogeneity for federated learning with adaptive client sampling. pages 1–10, 2022.

[4] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*, 2019.

[5] Hao Wang, Zakhary Kaplan, Di Niu, and Baochun Li. Optimizing federated learning on non-iid data with reinforcement learning. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*, pages 1698–1707. IEEE, 2020.

[6] Wenyu Zhang, Xiumin Wang, Pan Zhou, Weiwei Wu, and Xinglin Zhang. Client selection for federated learning with non-iid data in mobile edge computing. *IEEE Access*, 9:24462–24474, 2021.

[7] Yipeng Zhou, Yao Fu, Zhenxiao Luo, Miao Hu, Di Wu, Quan Z Sheng, and Shui Yu. The role of communication time in the convergence of federated edge learning. *IEEE Transactions on Vehicular Technology*, 2022.

[8] Jiawen Kang, Zehui Xiong, Dusit Niyato, Shengli Xie, and Junshan Zhang. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal*, 6(6):10700–10714, 2019.

[9] Jiawen Kang, Zehui Xiong, Dusit Niyato, Yuze Zou, Yang Zhang, and Mohsen Guizani. Reliable federated learning for mobile networks. *IEEE Wireless Communications*, 27(2):72–80, 2020.

[10] Tiansheng Huang, Weiwei Lin, Li Shen, Keqin Li, and Albert Y Zomaya. Stochastic client selection for federated learning with volatile clients. *IEEE Internet of Things Journal*, 2022.

[11] Takayuki Nishio and Ryo Yonetani. Client selection for federated learning with heterogeneous resources in mobile edge. In *IEEE international conference on communications (ICC)*, pages 1–7. IEEE, 2019.

[12] Jaewook Lee, Haneul Ko, and Sangheon Pack. Adaptive deadline determination for mobile device selection in federated learning. *IEEE Transactions on Vehicular Technology*, 2021.

[13] Sawsan AbdulRahman, Hanine Tout, Azzam Mourad, and Chamseddine Talhi. Fedmccs: Multicriteria client selection model for optimal iot federated learning. *IEEE Internet of Things Journal*, 8(6):4723–4735, 2020.

[14] Adnan Mahmood, Quan Z Sheng, Sarah Ali Siddiqui, Subhash Sagar, Wei Emma Zhang, Hajime Suzuki, and Wei Ni. When trust meets the internet of vehicles: opportunities, challenges, and future prospects. In *2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC)*, pages 60–67. IEEE, 2021.

[15] Adnan Mahmood, Sarah Ali Siddiqui, Quan Z Sheng, Wei Emma Zhang, Hajime Suzuki, and Wei Ni. Trust on wheels: Towards secure and resource efficient IoV networks. *Computing*, pages 1–22, 2022.

[16] Yining Liu, Keqiu Li, Yingwei Jin, Yong Zhang, and Wenyu Qu. A novel reputation computation model based on subjective logic for mobile ad hoc networks. *Future Generation Computer Systems*, 27(5):547–554, 2011.

[17] Yue Zou, Fei Shen, Feng Yan, Jing Lin, and Yunzhou Qiu. Reputation-based regional federated learning for knowledge trading in blockchain-enhanced IoV. In *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2021.

[18] Zhendong Song, Hongguang Sun, Howard H Yang, Xijun Wang, Yan Zhang, and Tony QS Quek. Reputation-based federated learning for secure wireless networks. *IEEE Internet of Things Journal*, 9(2):1212–1226, 2021.

[19] Jake Perazzone, Shiqiang Wang, Mingyue Ji, and Kevin Chan. Communication-efficient device scheduling for federated learning using stochastic optimization. pages 1–10, 2022.

[20] Hongda Wu and Ping Wang. Node selection toward faster convergence for federated learning on non-iid data. *IEEE Transactions on Network Science and Engineering*, 2022.

[21] Wenchao Xia, Tony QS Quek, Kun Guo, Wanli Wen, Howard H Yang, and Hongbo Zhu. Multi-armed bandit-based client scheduling for federated learning. *IEEE Transactions on Wireless Communications*, 19(11):7108–7123, 2020.

[22] Tiansheng Huang, Weiwei Lin, Wentai Wu, Ligang He, Keqin Li, and Albert Y Zomaya. An efficiency-boosting client selection scheme for federated learning with fairness guarantee. *IEEE Transactions on Parallel and Distributed Systems*, 32(7):1552–1564, 2020.

[23] Jiahua Ma, Xinghua Sun, Wenchao Xia, Xijun Wang, Xiang Chen, and Hongbo Zhu. Client selection based on label quantity information for federated learning. In *Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–6. IEEE, 2021.

[24] Jianxin Zhao, Xinyu Chang, Yanhao Feng, Chi Harold Liu, and Ningbo Liu. Participant selection for federated learning with heterogeneous data in intelligent transport system. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[25] Su Wang, Mengyuan Lee, Seyyedali Hosseinalipour, Roberto Morabito, Mung Chiang, and Christopher G Brinton. Device sampling for heterogeneous federated learning: Theory, algorithms, and implementation. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2021.

[26] Yann Fraboni, Richard Vidal, Laetitia Kameni, and Marco Lorenzi. Clustered sampling: Low-variance and improved representativity for clients selection in federated learning. In *International Conference on Machine Learning*, pages 3407–3416. PMLR, 2021.

[27] Zhiyuan Wang, Hongli Xu, Jianchun Liu, He Huang, Chunming Qiao, and Yangming Zhao. Resource-efficient federated learning with hierarchical aggregation in edge computing. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, pages 1–10. IEEE, 2021.

[28] Luca Barbieri, Stefano Savazzi, Mattia Brambilla, and Monica Nicoli. Decentralized federated learning for extended sensing in 6g connected vehicles. *Vehicular Communications*, 33:100396, 2022.

[29] Haibo Yang, Xin Zhang, Prashant Khanduri, and Jia Liu. Anarchic federated learning. In *International Conference on Machine Learning*, pages 25331–25363. PMLR, 2022.

[30] Dzmitry Huba, John Nguyen, Kshitiz Malik, Ruiyu Zhu, Mike Rabbat, Ashkan Yousefpour, Carole-Jean Wu, Hongyuan Zhan, Pavel Ustinov, Harish Srinivas, et al. Papaya: Practical, private, and scalable federated learning. *Proceedings of Machine Learning and Systems*, 4:814–832, 2022.