Oblivious Federated Analytics for Mobile Devices

Donghyun Sohn, Northwestern University, donghyun.sohn@u.northwestern.edu

As the volume of data generated from mobile devices surges, the need for efficient data analysis while preserving privacy is increasingly urgent. With federated analytics [3], edge devices perform distributed data analysis without participants uploading their records to a trusted third party. It presents a compelling solution to this quandary. For example, Google's Now Playing feature enables their servers to refine their song recommendation data by securely aggregating over the listening habits of individual users [5]. Here, a user's phone locally recognizes and records the songs they are hearing. It periodically encrypts their listening habits, and sends them to the a centralized server for analysis with those of other users. Although federated analytics offers efficient privacy-preserving data analysis, it inherently relies on centralized servers that learn the output of this analysis. This raises fundamental questions about trust: How can we ensure these servers do not misuse the aggregated data? Is it possible to create a system where users benefit from aggregated insights without the server ever "seeing" this data? The Fundamental Law of Information Recovery [1] indicates that revealing these raw statistics leaves users vulnerable to reconstruction attacks. We posit that this personalization is possible with better security by pushing more of this computation into mobile devices.

Secure multi-party computation (MPC) enables users to securely aggregate the union of their data without revealing anything to a trusted third party. To advance federated analytics in serverless architectures, we optimize MPC protocols by maximizing local computation. In our approach, devices start by aggregating their own data locally. Participants next secret share these aggregates, getting them into an encrypted form with which they may jointly compute over them securely with the data of their peers. This transformation involves splitting the aggregated data into pieces that are meaningless on their own but can be combined to reveal a specific piece of information. These shares are distributed among the participating devices, initiating the MPC process. Within this MPC framework, a series of cryptographic protocols are executed, allowing the devices to collaboratively compute the final aggregate without any single device having to reveal its own aggregated data to others. Once the final aggregation is complete, the result is revealed to participating devices. Our methodology is expected to facilitate privacy-preserving analysis across multiple parties without revealing individual data.

Furthermore, we are tailoring MPC protocols to manage a continuous querying environment effectively, distinguishing our work from existing models, e.g., SMPAI [4]. Recognizing the limited resources of edge devices [2], our approach optimizes partial aggregation through adaptive sampling and incremental aggregation techniques. This allows for selective querying based on data relevance and likelihood of change, substantially reducing local computational overhead.

By integrating MPC and local data aggregation within federated analytics for edge devices in a continuous querying environment, our approach contributes to introducing a federated analytics model that eliminates the need for server visibility into aggregated data. Our experimental evaluation will focus on measuring computational overhead, latency, and bandwidth usage during the secure aggregation process, anticipating significant enhancements in computational efficiency. This effort opens pathways for secure and efficient federated analytics across edge devices. Furthermore, by leveraging aggregated data to improve personal models without server knowledge, it promotes private device-level personalization.

[1] John Abowd, Lorenzo Alvisi, Cynthia Dwork, et al. Privacy-preserving data analysis for the federal statistical agencies. *arXiv preprint arXiv:1701.00752*, 2017.

[2] Keyan Cao, Yefan Liu, Gongjie Meng, and Qimeng Sun. An overview on edge computing research. *IEEE Access*, 8:85714–85728, May 2020.

[3] Ahmed Ramzy Elkordy, Youssef H Ezzeldin, and Shuguang Han. Federated analytics: A survey. *IEEE Transactions on Signal and Information Processing over Networks*, 2023.

[4] Vaikkunth Mugunthan, Antigoni Polychroniadou, David Byrd, and Tucker Hybinette Balch. Smpai: Secure multi-party computation for federated learning. In *Proceedings of the NeurIPS 2019 Workshop on Robust AI in Financial Services*. jpmorgan.com, 2019.

[5] Daniel Ramage and Stefano Mazzocchi. Federated analytics: Collaborative data science without data collection. https://blog.research.google/2020/05/federated-analytics-collaborative-data.html, 2020.