**Meenakshi Nagarajan**

**RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response**

**Research problem and its importance**

Knowledge about user’s experience on a product, usually collected through surveys, is highly essential to make improvements to any product or a service. The collected data will be analyzed statistically to better learn about the population and make informed decisions on product’s improvement. As this technique has become increasingly popular, user participation in the surveys increases and they tend to release more information about them. Frequent and multiple statistical analysis on the data obtained, permits the survey operator to gain access to confidential information about the user. To overcome the privacy risk and to allow high utilization of collected data, better privacy mechanisms needs to be introduced.

**Technical challenges of the problem**

* As mentioned in the problem statement, because of privacy concerns collecting useful data from the users for analysis is a great challenge. Therefore, the data collection design must be handled carefully.
* The person who collects the data should make sure that there are no linkable identifiers that can conclude any sensitive information about the user
* Data collection should not depend on any untrustworthy cloud services where there are high chances of privacy risk.

**Proposed approach**

RAPPOR uses Bloom filter based randomized response technique on strings. Once the data has been collected in the bloom filter, two levels of noise will be introduced on the data. The first step is Permanent randomized response that is used to create a noisy answer. It also uses memoization technique to memoize the noisy answer unique to an individual and to reuse it in place of the real answer. The second step is Instantaneous randomized response, to report the result of noisy answer. These results are then sent to the server. The large amount of randomness in the response makes sure that the operator cannot draw meaningful conclusions about the user.

**Strengths of the approach**

* No impact from privacy externalities
* Protects the users’ privacy along with high utility of data while the privacy doesn’t depend on a trusted third party
* Memoization can allow the service operator to obtain multiple responses from the same user without affecting their privacy
* There is no central database to store the user data which are prone to attacks

**Possible weaknesses**

* Users participating in a survey more than once from different accounts resulting in multiple Permanent randomized responses could release more information and are prone to privacy attacks
* Complex statistical methods might be needed to deduce results
* As the memoization technique reuse the randomized bits for all the responses from a user, it could become a unique tracking id and can affect client’s anonymity