SEREN: A DIFFERENTIAL PRIVACY APPLICATION FOR NETWORKING.

George Igwegbe*, Lawrence Francis*, Orevaoghene Ahia*, Kelechi Ogueji*, Azeez Oluwafemi*, Amel Sellami, Marek Barwinski, Tejumade Afonja**

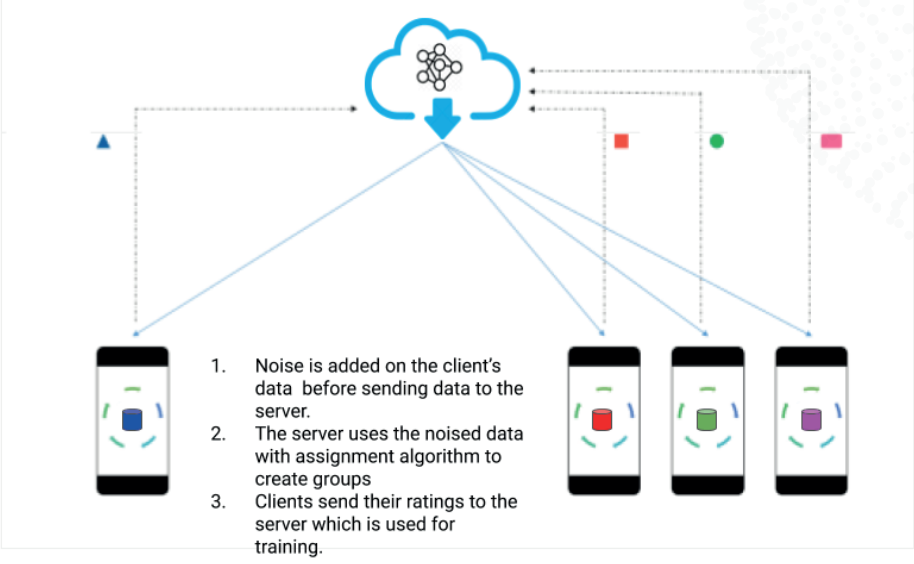
InstaDeep™

Introduction

Conference attendees are frequently overwhelmed by the scale of the event they participate in. Young participants, most especially find it difficult to identify the right co-attendees to reach out and connect with. We believe that by creating a simple and intuitive app that removes the anxiety and uncertainty from arrangement of such meetings we will dramatically increase the quantity and quality of connections. Following each match-making we collect user feedback and retrain our model to improve the quality of recommendations before the next round. The goal is for users to report that through the usage of this app they were able to make more meaningful connections during the conference than they would otherwise be able to. We used a differential privacy approach in order to protect users' data in training the model as opposed to a regular machine learning approach.

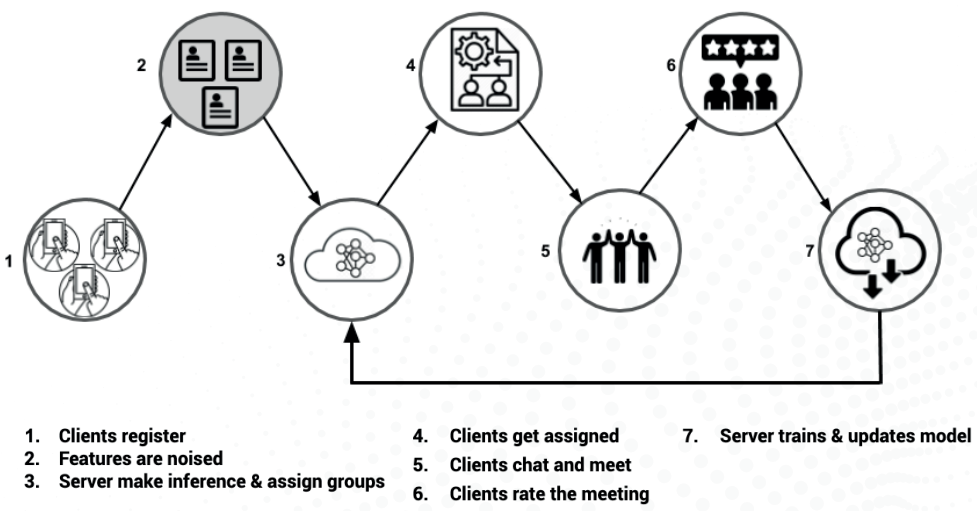
Software Architecture

In order to protect users' data, we implemented local differential privacy on users' devices before sending their data to the server. We applied the algorithms shown in our methodology based on Wang et al [1].



Methodology

1. Group assignment: Each user is assigned to group using Algorithm 1.
2. The user attends the meeting and give a numerical rating between 1 and 5.
3. The user's rating is sent to the server and training is done on the server.
4. We repeat for the next meeting round.



References

1. Ning Wang, Xiaokui Xiao, Yin Yang, Jun Zhao, Siu Cheung Hui, Hyejin Shin, Junbum Shin, and Ge Yu, "Collecting and Analyzing Multidimensional Data with Local Differential Privacy", arXiv preprint arXiv:1907.00782
2. Carnegie Mellon University (2011). StatLib---Datasets Archive. Retrieved April 21, 2011 from http://lib.stat.cmu.edu/datasets/boston
3. Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani (2004) (Least Angle Regression) Annals of Statistics (with discussion), 407-499

Algorithm 1: Group Assignment Algorithm

n = number of users in a group.
N = total number of users.
U = set of all users, $\{u_1, u_2, \ldots, u_N\}$.

1. Randomly pick a user, u_i from **U**.
2. Continuously take a random sample sets of **(n-1)** other users from **U** without replacement.
3. For each set of **(n - 1)** other users form a group feature vector for user ui and others and use a model to predict group rating.
4. Assign user u_i to the group with the highest rating.
5. Repeat steps 1 to 4 till all users are assigned to a group.

Local Differential Privacy (LDP)

LDP is a recently proposed privacy standard for collecting and analyzing data to guarantee that sensitive information of an individual cannot be inferred with high confidence.

Wang et al's algorithm 2 & 3 explains the process of perturbing the multidimensional users' features.

Algorithm 2: Wang et al.'s Solution [1] for Multiple Numeric Attributes.

input: tuple $t_i \in [-1,1]^d$ and privacy parameter ϵ.
output: tuple $t^*_i \in [-C \cdot d, C \cdot d]^d$.

1. Let $t^*_i = (0,0,\ldots,0)$;
2. Let k = max{1, min {d,}};
3. Sample k values uniformly without replacement from {1,2, …., d};
4. for each sampled value j do
5. Feed ti [Aj] and ϵ/k as input to PM and obtain a noisy value x_i, j;
6. $t^*_i[A] = (d/k) * x_i$,j ;
7. return t^*_i

Algorithm 3: Piecewise Mechanism for One-Dimensional Numeric Data. [1]

input: tuple $t_i \in [-1,1]$ and privacy parameter ϵ.
output: tuple $t^*_i \in [-C, C]$.

1. Sample x uniformly at random from [0,1];
2. If $x < e^{\epsilon/2}/(e^{\epsilon/2}+1)$ then
3. sample t_i^* uniformly at random from $[\ell(t_i), r(t_i)]$;
4. **else**
5. sample t_i^* uniformly at random from $[-C, \ell(ti)) \cup (r(ti), C]$
6. **return** t_i^*

Experimental Evaluation and Result

In order to verify the LDP algorithm, we trained and evaluated on two regression datasets. The Boston House Prices [2] and Diabetes Progression [3]. Our method (middle column) preserves privacy with minimal impact on error rate.

Dataset	MSE on original dataset	MSE on privacy preserving dataset	MSE with untrained model (Baseline)
Boston	0.0700	0.0746	5.0647
Diabetes	0.0754	0.1116	0.5264

Conclusion

We successfully built a conference networking app with users' data protected using a local differential privacy method based on Wang et. al [1] and also verified their method by demonstrating they performed better than random albeit with an accuracy tradeoff. In the future, we plan to solve a similar problem by clustering in a federated learning setting. The advantage would be a performance gain in addition to user privacy.

*** these authors contributed equally to the work ** corresponding author.** **We would like to acknowledge Abiola Lapite and Zohra Slim for their valuable contribution to this work.**