New Algorithms for Differential Privacy

Mary Scott (University of Warwick), supervised by Graham Cormode and Carsten Maple

Modern data analysis relies on gathering data from individuals that is considered highly sensitive – for example their medical history. There is a pressing need to develop trustworthy systems that limit the exposure of this sensitive information. One possible approach to this problem is to allow data disclosure whilst protecting privacy via anonymisation. A more rigorous notion of privacy introduced in 2006 is Differential Privacy (DP), which guarantees that the output of a computation on a dataset is not changed significantly upon the removal of any individual from that dataset. Local Differential Privacy (LDP), a model of DP with an additional restriction, is used by hundreds of millions of people every day. My future research will focus on extending the recently developed Single-Message Shuffle Model (SMSM).

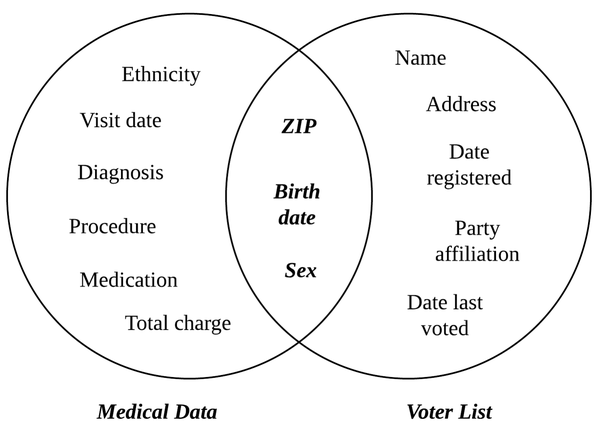

Figure 1: Linking to re-identify data.

Background

Last summer, I completed an internship exploring the issue of *linkage*, which occurs when an adversary finds two datasets with sufficient common information about the same people that they can be merged. Refer to Figure 1 for an example.

I primarily studied KHyperLogLog (KHLL), an algorithm that could estimate the *re-identifiability* (the probability of recovering the identities of people) and the *joinability* (the probability of finding common datasets were linkable by unexpected join keys) risks of very large databases.

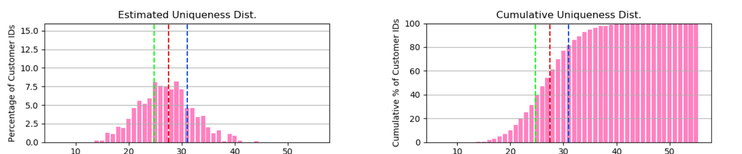

Figure 2: Re-identifiability graphs for $K = 100$.

I wrote a program in Python for KHLL that satisfied linearity and required minimal memory, concluding that the risk of linkage decreased as the random sample size K increased, and increased as the intersection I increased. See Figure 2 for an example output of my program.

My research on KHLL relied on anonymising datasets, to allow data disclosure without violating privacy. I have now moved on to a different approach: to actually control the disclosure of data.

National Cyber Security Centre
a part of GCHQ

> "Differential Privacy formally defines what it means for a computation to be privacy-preserving"
>
> **Narayanan and Shmatikov**
> The researchers who de-anonymised the Netflix Prize dataset

The ε-Differential Privacy Guarantee

One of the fundamental challenges of data analysis is the careful balance of acquiring as much utility from a dataset as possible, whilst simultaneously providing a strong guarantee of privacy to each individual affected.

A function applied to a dataset is *differentially private (DP)* if, with the removal of any individual from the dataset, the output of the function does not change more than a small multiplicative factor ε.

Definition [1]: A randomised function M gives *ε-differential privacy* if for all datasets D_1 and D_2 differing on at most one element, and all $S \subseteq Range(M)$,

$$\Pr[M(D_1) \in S] \leq \exp(\varepsilon) \times \Pr[M(D_2) \in S].$$

Randomised Response

Imagine an organisation is collecting data about the voting habits of a sample of students studying at the University of Warwick. If a representative were to directly ask each participant "Did you vote for party P in the last general election?" then this violates their privacy.

In *randomised response*, participants report whether or not they voted for a party P by flipping a coin:

- If the coin lands tails, then they respond truthfully;
- If the coin lands heads, then they flip a second coin and respond "Yes" if heads and "No" if tails.

To estimate the true fraction p of participants who voted for party P, start by calculating the expected number E of "Yes" answers as seen in [2]:

$$E = \frac{1}{4}(1-p) + \frac{3}{4}p = \frac{1}{4} + \frac{p}{2}.$$

Then rearrange the expression above to estimate p as $2E - \frac{1}{2}$. Randomised Response can be linked to DP: the parameter ε can be expressed as a function of p.

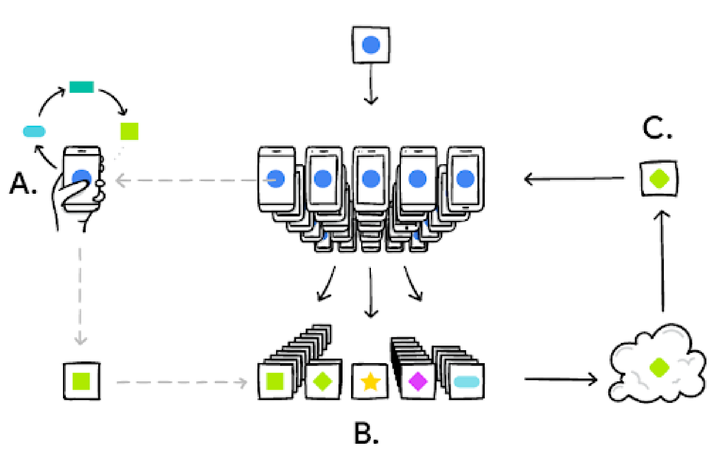

Figure 3: A user's phone personalises the model locally, based on their usage (A). Many users' updates are aggregated (B) to form a consensus change (C) to the shared model, after which the procedure is repeated.

Local Differential Privacy

A more recent development is Local Differential Privacy (LDP), which requires that the output of *every* user meets the ε-differential privacy guarantee. This guarantee is fulfilled at the local stage, rather than after the data is collected, meaning that the personal information of every user remains private even from data analysts. See Figure 3 for a visual explanation.

LDP is a superior form of privacy as it does not require a trusted third party. It has been deployed at large scale in several of the world's most popular companies, including Google [3] and Apple [4].

Apple's LDP implementation used the Count Mean Sketch to collect emoji usage data, with results shown in Figure 4. This helped the company design better ways for their consumers to find and use their favourite emojis.

Note that Randomised Response gives an LDP result: the mechanism protects the privacy of any specific participant, irrespective of any attacker's prior knowledge.

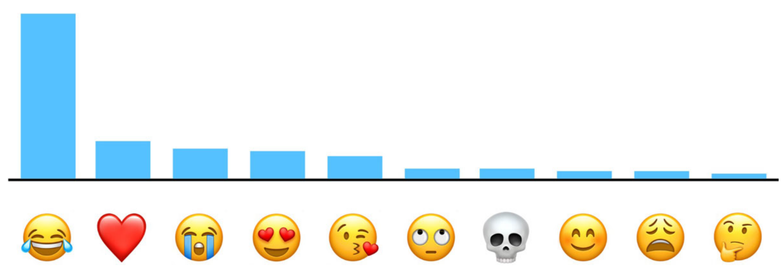

Figure 4: The most popular emojis for Apple customers, collected using Local Differential Privacy.

The Single-Message Shuffle Model

In the Single-Message Shuffle Model (SMSM), the data collector receives one message from each of the users as in LDP. The crucial difference is that SMSM assumes the data collector is unable to associate messages to users.

SMSM differs significantly from LDP in terms of *assumed trust*: SMSM requires users to provide messages carefully crafted to protect each other's privacy, as well as relying on a trusted shuffling step. This is in contrast with DP where the responsibility is entirely held by the trusted curator.

In [5], Balle developed a single-message shuffle protocol for the private summation of (real) numbers $x_i \in [0, 1]$. I plan to extend his argument to include the private summation of vectors.

References

[1] C. Dwork. Differential privacy. *Automata, Languages and Programming*, pages 1–12, 2006.

[2] C. Dwork and A. Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

[3] Ú. Erlingsson. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067. ACM, 2014.

[4] ADP Team. Learning with privacy at scale. *Apple Mach. Learn. J*, 1(9), 2017.

[5] B. Balle. The privacy blanket of the shuffle model. *arXiv preprint arXiv:1903.02837*, 2019.