Differential Privacy

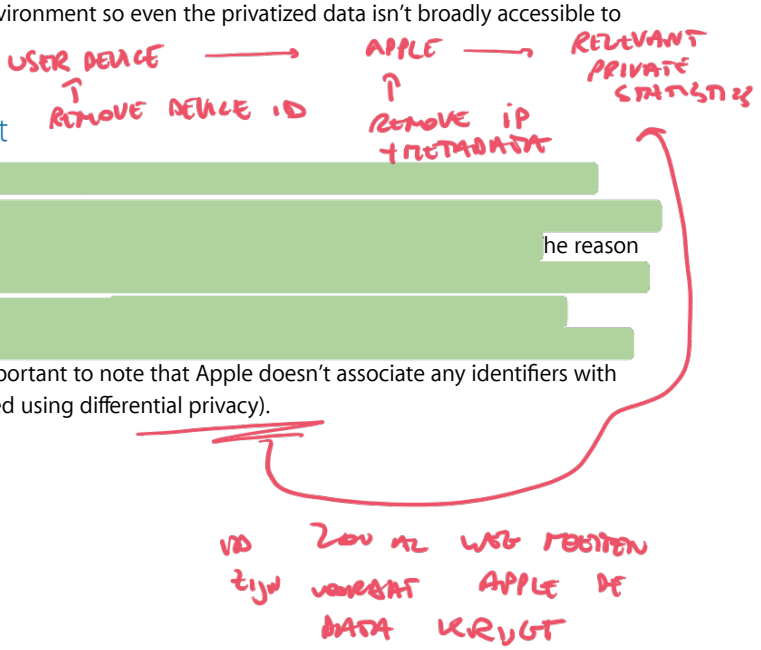
There are situations where can for example: What new words are trending and might make the most relevant suggestions? What websites have problems that could affect battery life? Which emoji are chosen most often? The challenge is that the data which could drive the —such as what the users type on their keyboards—

A privacy-preserving system

Apple has adopted and further developed a technique known in the academic world as to do something really exciting It is a technique that enables Apple to learn about the user community without learning about individuals in the community

The differential privacy technology used by Apple is and Apple can see meaningful information emerge.

. The system is and designed to provide The first step we take is to privatize the information using local differential privacy on the user's device. The purpose of privatization is to assure that The Apple analysis system ingests the differentially private contributions, dropping IP addresses and other metadata. The final stage is aggregation, where the privatized records are processed to compute the relevant statistics and the aggregate statistics are then shared with relevant Apple teams. Both the ingestion and aggregation stages are performed in a restricted access environment so even the privatized data isn't broadly accessible to Apple employees.



Privacy budget

The he reason is that the (though it's important to note that Apple doesn't associate any identifiers with information collected using differential privacy).

Apple uses local differential privacy to help protect the privacy of user activity in a given time period, while still gaining insight that improves the intelligence and usability of such features as:

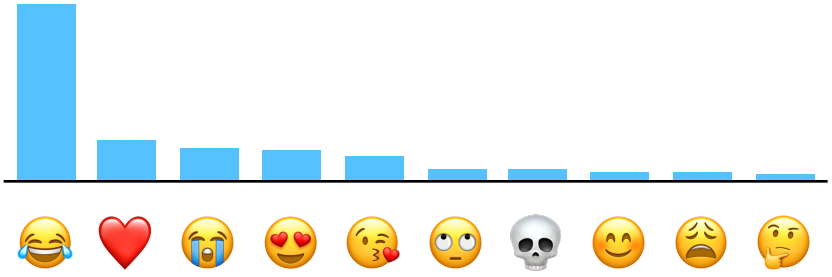
- QuickType suggestions
- Emoji suggestions
- Lookup Hints
- Safari Energy Draining Domains
- Safari Autoplay Intent Detection (macOS High Sierra)
- Safari Crashing Domains (iOS 11)
- Health Type Usage (iOS 10.2)

For each feature, Apple seeks to make the privacy budget small while still collecting enough data to to enable Apple to improve features.

For Lookup Hints, Apple uses a privacy budget with epsilon of 4, and limits user contributions to two donations per day. For emoji, Apple uses a privacy budget with epsilon of 4, and submits one donation per day. For QuickType, Apple uses a privacy budget with epsilon of 8, and submits two donations per day.

For Health types, Apple uses a privacy budget with epsilon of 2 and limits user contributions to one donation per day. The donations do not include health information itself, but rather which health data types are being edited by users.

For Safari, Apple limits user contributions to 2 donations per day. For Safari domains identified as causing high energy use or crashes, Apple uses a single privacy budget with epsilon of 4. For Safari Auto-play intent detection, Apple uses a privacy budget with epsilon of 8.



The Count Mean Sketch technique allows Apple to determine the most popular emoji to help design better ways to find and use our favorite emoji. The top emoji for US English speakers contained some surprising favorites.

Techniques

Local differential privacy guarantees that it is difficult to determine whether a certain user contributed to the computation of an aggregate by adding slightly biased noise to the data that is shared with Apple.

Apple currently makes use of

In our use of the Count Mean Sketch technique for differential privacy, the original

The data is encoded using variations of a SHA-256 hash followed by a privatization step and then written into the sketch matrix with its values initialized to zero.

The noise injection step works as follows: where This assures that analysis of the collected data cannot distinguish actual values from flipped values, helping to assure the privacy of the shared information.

When the information encoded in the sketch matrix is sent to Apple,

Hadamard Count Mean Sketch

The Hadamard Count Mean–based Sketch technique uses a noise injection method but with an important difference:

Seeing user data

Users can examine the information being shared with Apple for the categories of data that are protected using Differential Privacy. In iOS, the information is visible under Settings > Privacy > Analytics > Analytics Data, in entries that begin with "DifferentialPrivacy." In macOS, users can launch the Console app and view the information under the Differential Privacy category of System Reports.

Controlling participation

The data-gathering features that use differential privacy are linked to the user setting for Device Analytics. Users are presented with the option of sending diagnostic information when they set up a device running macOS or iOS, and they can always change their choice later in System Preferences on macOS or the Settings app on iOS.

The beginning

Apple launched differential privacy for the first time in macOS Sierra and iOS 10. Since then, we have expanded to other use cases such as Safari and Health types. As Apple continues to refine differential privacy algorithms, we look forward to using them to improve user experience in other areas of our products, while continuing to work to protect our users' private information.