

PPM & Differential Privacy

Christopher Patton

PPM - IETF 115 - London

Idea for draft: Differential Privacy Guidelines for PPM

Motivation: Limitations of DAP

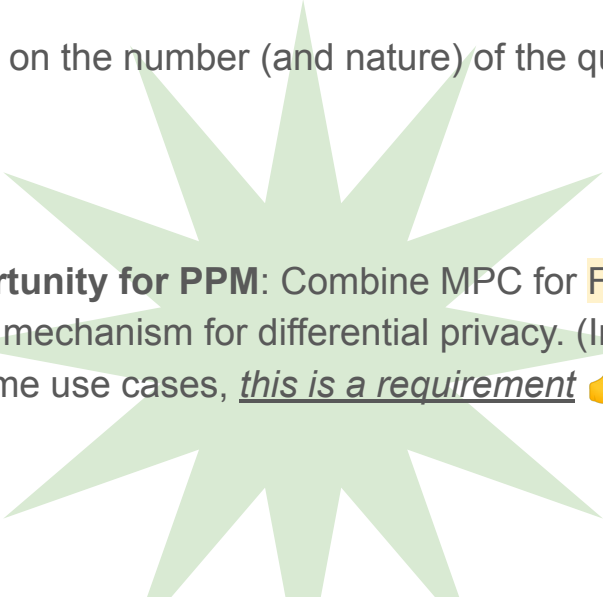
- DAP provides MPC*-style security guarantees: Collector learns some aggregation function $F(DB)$ of a batch of measurements $DB = m[1], \dots, m[n]$ and nothing else
- While **necessary** in our threat model, MPC is **not sufficient** for privacy
 - Canonical example: Over-sampling a user by including multiple measurements in a single batch, or across multiple batches [1]

[1] <https://datatracker.ietf.org/meeting/114/materials/slides-114-ppm-collect-sub-protocol-privacy-requirements>

* MPC = "Multi-Party Computation"

Overview of Differential Privacy

- Dwork '06 [1]: A batch query algorithm **Query** is **differentially private** if the distribution of **Query(DB₁)** is "close to" the distribution of **Query(DB₂)** for all batches **DB₁**, **DB₂** differing in exactly one measurement
 - Example: Aggregate the measurements into **F(DB)**; sample **noise N** from a "suitable" distribution; and return **F(DB) + N**
 - **Privacy budget**: Degree of privacy depends on the number (and nature) of the queries



Opportunity for PPM: Combine MPC for **F(DB)** with a mechanism for differential privacy. (In fact, for some use cases, this is a requirement 👍)

OK, but how?

- Lots of public discussion on this point:
 - DAP issues 19, 20, 210 [1]
 - VDAF issue issue 94 [2]
 - List [3]
- Key takeaways:
 - Most use cases PPM aims to address can benefit from differential privacy.
 - Implementing differential privacy correctly hard.
 - The most suitable mechanism depends not only on the base protocol (e.g., STAR or DAP/Poplar1) but also the nature of the measurements and how they're used.
 - **No concrete proposals, yet.**

[2] <https://github.com/ietf-wg-ppm/draft-ietf-ppm-dap/issues>

[3] <https://github.com/cfrg/draft-irtf-cfrg-vdaf/issues>

[4] https://mailarchive.ietf.org/arch/msg/ppm/2d4aPwkSlvczXgjj_LwG7HreP68/

Open questions (if time)

- What is in scope for the draft?
 - Algorithms for sampling noise (where applicable)
 - Enforcing privacy budget (think "safety margins")
 - (Un)suitable applications
- Should the draft specify concrete mechanisms?
 - DAP/Prio3CountVec with local DP described in [4, Section 4.2]
 - DAP/Poplar1 with the central DP described in [5, Appendix E]

[5] Google and Apple, "ENPA White Paper", 2021.

[6] Boneh et al., "Lightweight Techniques for Private Heavy Hitters", S&P 2021.