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 $DB = m[1], ..., m[n]$ 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]


* 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

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.**

---

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]