Randomized Requantization with Local Differential Privacy

Sijie Xiong¹, Anand Sarwate¹, Narayan Mandayam¹

¹ Rutgers, The State University of New Jersey

Why interesting?

- Goals for future sensor networks such as IoT:
  - limited resource utilization
  - maintain data fidelity
  - protect private information
  - avoid attacks on individual nodes

What are the tradeoffs between these criteria?

The system model

Each sensor $s_i$ transmits a value $X_i$ or $Y_i$, which the server combines to compute the aggregate signal $X$. The server does not know the distribution $D$ of $X_i$.

Performance metrics

- Local differential privacy [Duchi et al. '13]:
  - The adversary’s likelihood of guessing that the input was $x$ over $x'$ doesn’t increase more than $e^\epsilon$ after observing the released value $y$.
  - $P(x = x) = P(x = x'|y = y) \cdot e^\epsilon$

- Compression ratio: $\rho = \log_2|Y| / \log_2|X|$

- Utility (mse):
  - $\delta = E_{P,Q}[d(X,Y)]$
  - $\sum_{i=1}^{N} \sum_{j=1}^{N} P(x_i)Q(y_j|x_i)(x_i - y_j)^2$

Goal: find privacy-utility tradeoff and optimal $Q$

- The set of $\epsilon$-locally differentially private channels and the set of channels yielding expected distortion no greater than $\delta$ are defined by
  - $Q_{LDP}(\epsilon) = \left\{ Q(y|x) : \log_{\frac{Q(y|x)}{Q(y|x')}} \leq \epsilon, \forall (x, x', y) \in X \times X' \times Y \right\}$
  - $Q_{MSE}(\delta) = \left\{ Q(y|x) : \max_{P \in P} E_{P \times Q}[d(X,Y)] \leq \delta \right\}$

- Given $P$, $\rho$, $\delta$, the optimal $\epsilon$ becomes
  - $\epsilon^* = \min_{\epsilon \geq 0}\max_{P \in P} E_{P \times Q}[d(X,Y)] \leq \delta$

- S. under $P \times Q$

Theorem

The above optimization problem is a constrained quasi-convex optimization problem, and can be solved by bisection method.

Solving the optimization problem

Minimum achievable privacy level $\epsilon^*$ given $\delta, \rho$ value pairs, finding

- $(\epsilon, \delta, \rho)$-tradeoff
  - for fixed $\rho$, standard $\delta \uparrow \epsilon$ tradeoff
  - across cmp. ratios, achievable $\epsilon$ quite small under small $\delta$
  - can halve bit rate without sacrificing privacy

Validation on synthetic data

- Compare randomized requantization (RR) with perturbation method in the sparse Fourier transform domain
  - RR works better, more consistent
  - RR adds in much smaller noise
  - RR scales better with network size

Ongoing work and further directions

- Optimizing over reconstruction $Y$ (c.f. Lloyd-Max).
- Use privacy allocation to apportion resources in networks:
  - Individuals have different privacy budget $e_1, e_2, \ldots, e_N$
  - Multiple servers trying to access the same data
  - Gateway has to manage constraints and demands

Diagram:

- Sensors $X_1, X_2, \ldots, X_N$ transmit to a gateway, which passes on the information to a server.