Real-Time Aggregate Monitoring under Differential Privacy

Liyue Fan, Li Xiong
Department of Math & Computer Science
Emory University
Real-Time Aggregate

- **Disease Surveillance**
  - E.g. daily count of flu cases at a hospital
- **Traffic Monitoring**
  - E.g. hourly count of vehicles at a highway junction

**Goal:**
Strong Privacy, High Utility
Differential Privacy [BLR08]

Privacy Budget

\[ \Pr[A(D) \in S] \leq e^\alpha \times \Pr[A(D') \in S] \]

Randomized Algorithm

Neighboring Databases: differ in exactly one entry

Any Measurable Set

\[ A(D) = f(D) + \text{Lap}(\Delta f) \]

\[ \Delta f = \max_{D,D'} \| f(D) - f(D') \|_1 \]

Function Sensitivity

\[ \text{Lap}(\lambda) \sim \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} \]

e.g. \( \Delta \text{count} = 1 \)
Problem Statement

• A univariate, discrete Time-Series $X = \{x_k\}$ with $0 \leq k < T$

• **Problem**: Given time series $X$ and differential privacy budget $\alpha$, release $\alpha$-differentially private series $R$ with high utility.

• Utility: relative error
Challenges

- High sensitivity - T
- Low utility - Lap(T/α)
- Real-time requirement

• **Existing methods:**
  - Baseline LPA
    - Applies Laplace perturbation at every time stamp
    - Low Utility
  - State-of-the-art DFT
    - Performs Discrete Fourier Transform to the raw aggregate series
    - Reduced sensitivity, not applicable to real-time applications

• **Sampling**
• **Model-based Estimation**
• **Feedback**
FAST: a real-time system with Filtering and Adaptive Sampling for monitoring aggregate Time-series

Diagram:
- Time-series
  - Sampling point
  - Laplace Perturbation
    - Prediction
    - Correction
- Adaptive Sampling
  - error
- output
  - Sampling rate
Filtering

- Process Model
  
  \[ x_{k+1} = x_k + \omega \]
  \[ \omega \sim \mathcal{N}(0, Q) \]
  
  *Process noise*

- Measurement Model
  
  \[ z_k = x_k + \nu \]
  \[ \nu \sim \text{Lap}(\lambda) \]
  
  *Measurement noise*

- Approximate measurement noise with Gaussian
  
  \[ \nu \sim \mathcal{N}(0, R) \]

→ the Kalman filter
Sampling

- **Fixed-Rate Sampling**
  - Periodically sample the time series
  - Difficult to determine optimal sampling interval a priori

- **Adaptive Sampling**
  - Adjust the sampling rate/interval based on feedback
  - Implemented by PID control
  - Error to measure the performance of the sampling process
Evaluation: Data Sets

- Flu: CDC/flu, 209 data points
- Traffic: UW/intelligent transportation systems research, 540 data points
- Unemployment: St. Louis Federal Reserve Bank, 478 data points
Utility vs. Privacy

Flu

Traffic

Unemployment

Average relative error vs. $\alpha$ for different methods: FAST, LPA, DFT.
Conclusion

• **Contributions:**
  • Establish the *state-space* model for real-time aggregate under *differential privacy*
  • *Adaptively sample* the data series to reduce perturbation noise
  • Dynamically adjust the sampling rate and estimation based on *feedback*
  • Demonstrate the superior performance of FAST with real-world data sets

• **On-going Work:**
  • Accurate posterior estimation
  • Extension to sharing spatio-temporal data sets

• **Questions?**
  • Contact: liyue.fan@emory.edu
  • AIMS Group: www.mathcs.emory.edu/aims