

CIKM'12

Real-Time Aggregate Monitoring under Differential Privacy

Liyue Fan, Li Xiong

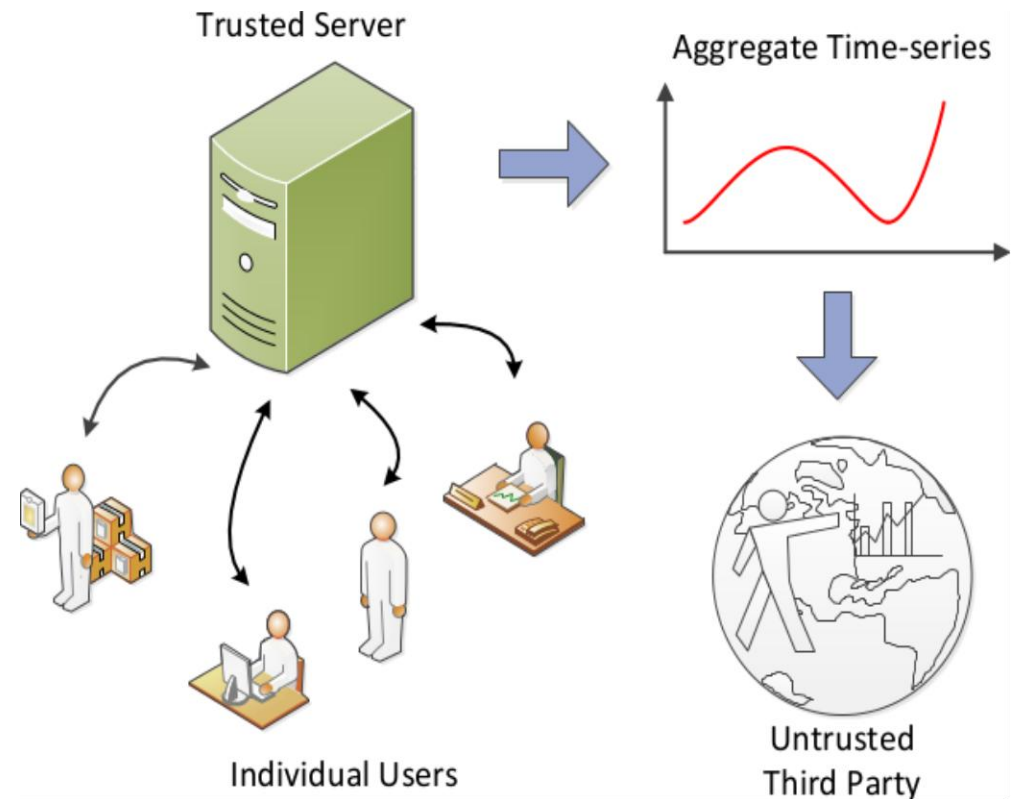
Department of Math & Computer Science

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

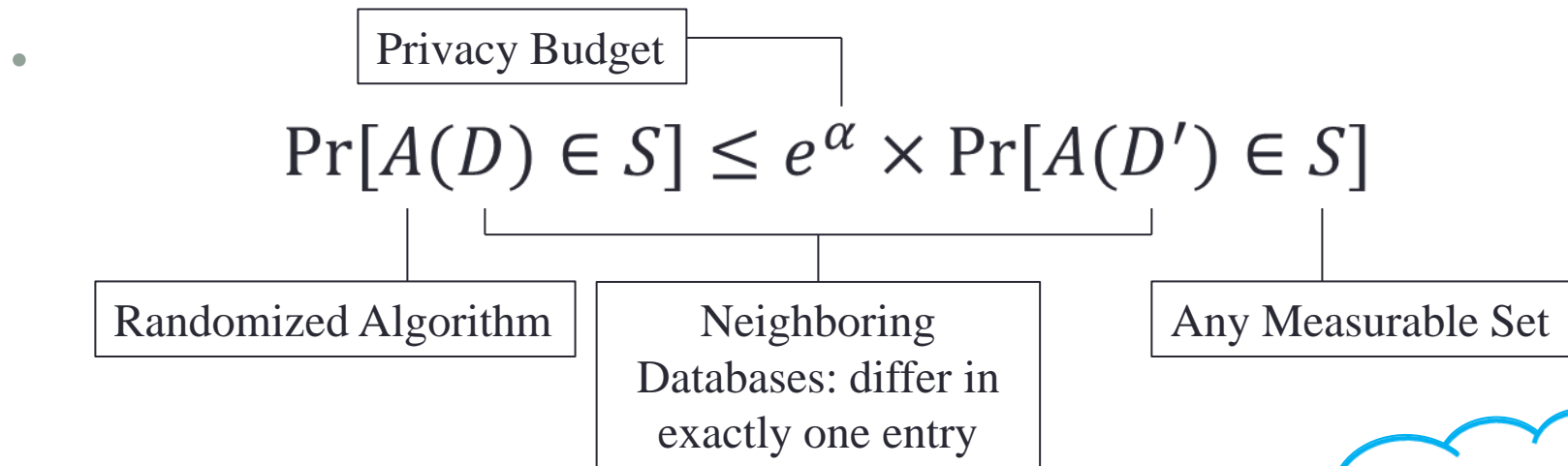


Goal:

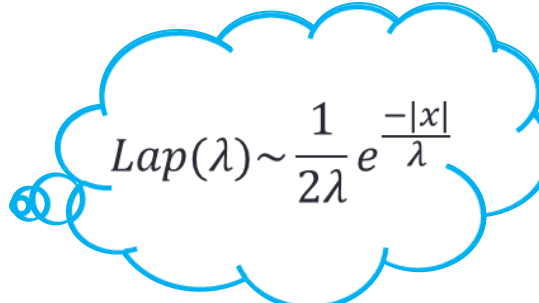
Strong Privacy, High Utility



Differential Privacy [BLR08]



$$A(D) = f(D) + \text{Lap}\left(\frac{\Delta f}{\alpha}\right)^d$$



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

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

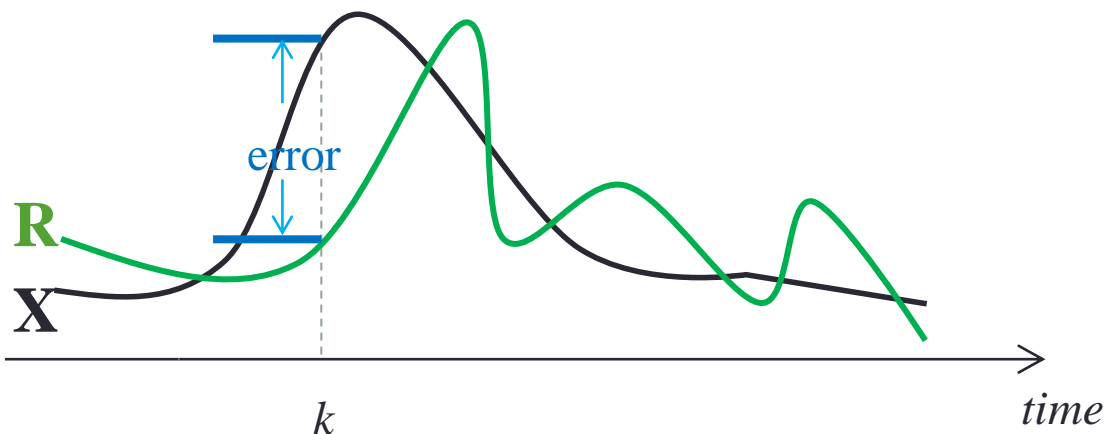
e.g. $\Delta \text{count} = 1$

Function Sensitivity



Problem Statement

- A univariate, discrete *Time-Series* $\mathbf{X} = \{x_k\}$ with $0 \leq k < T$
- **Problem:** Given time series \mathbf{X} and differential privacy budget α , release α -differentially **private** series \mathbf{R} with **high utility**.
- Utility: relative **error**



Challenges

- **High** sensitivity - T
- **Low** utility - $\text{Lap}(T/\alpha)$
- **Real-time** requirement



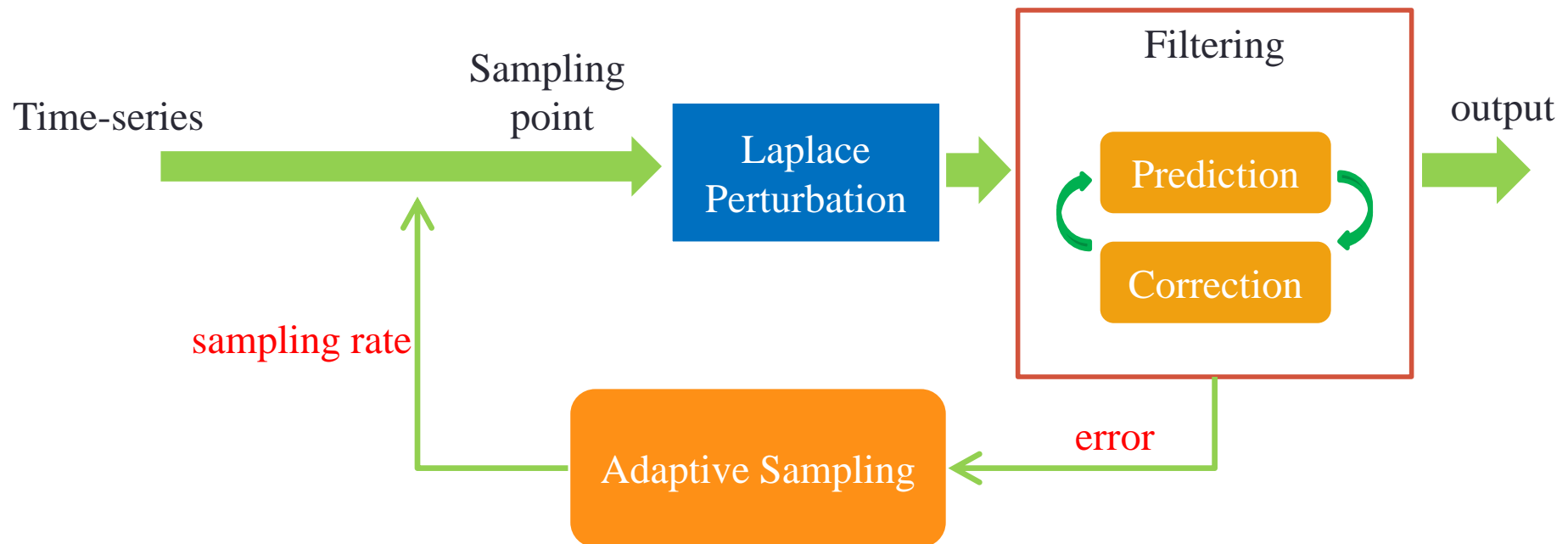
- **Existing methods:**

- **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



Filtering

- Process Model

$$x_{k+1} = x_k + \omega$$
$$\omega \sim \mathcal{N}(0, Q)$$

Process noise

- Measurement Model

$$z_k = x_k + v$$
$$v \sim \text{Lap}(\lambda)$$

Measurement noise

- Approximate measurement noise with Gaussian

$$v \sim \mathcal{N}(0, R)$$

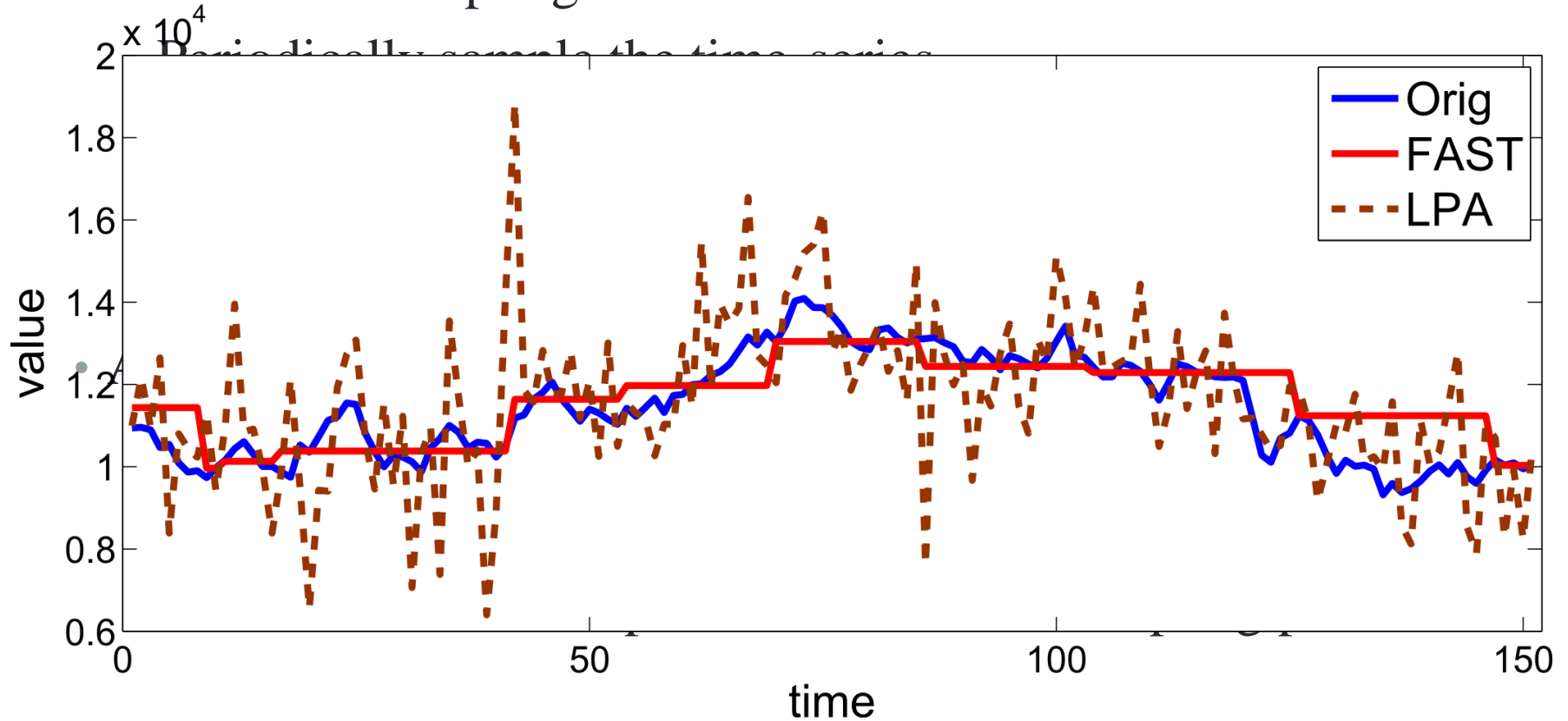
→ the Kalman filter



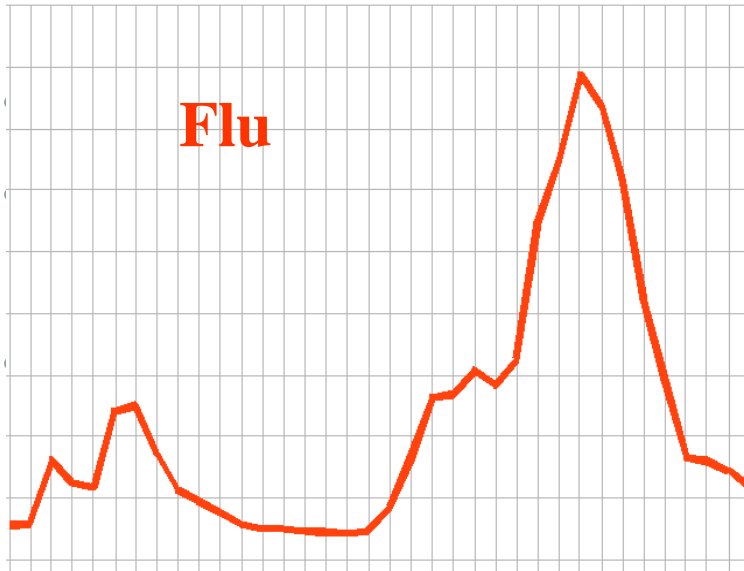
Sampling

- Fixed-Rate Sampling

Periodically sample the time series



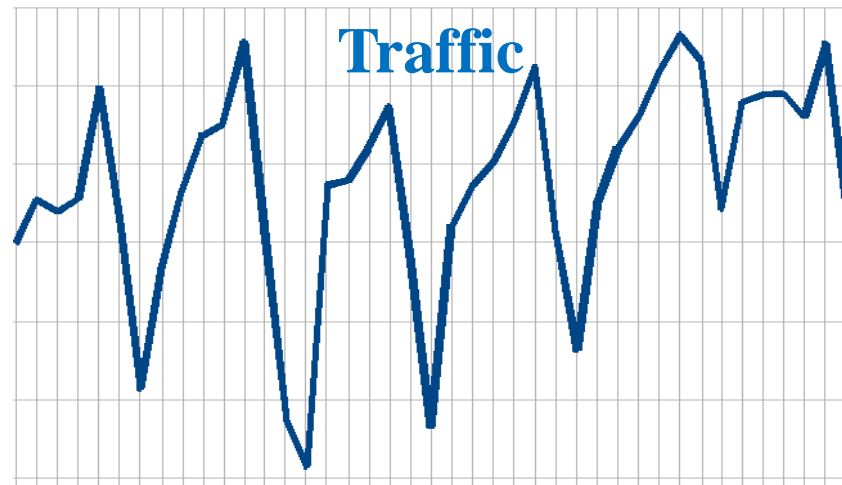
Evaluation: Data Sets



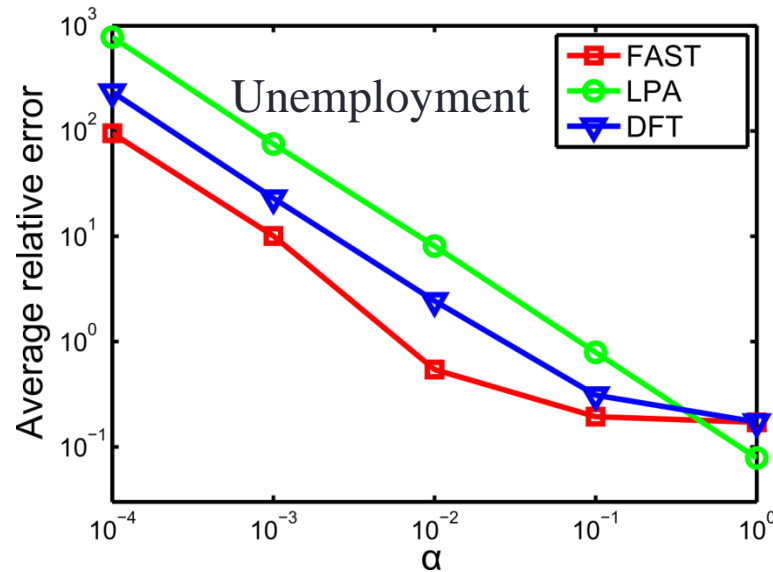
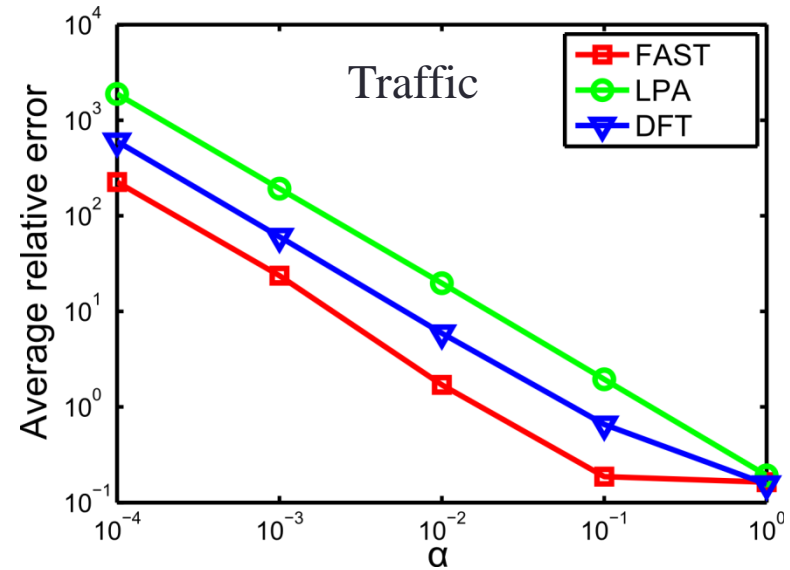
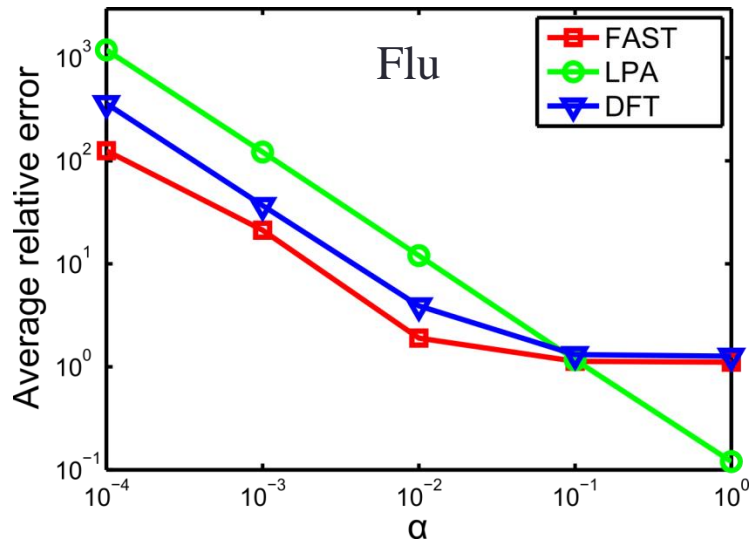
ts

transportation systems research, 540

Federal Reserve Bank, 478



Utility vs. Privacy



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

