PoolView: Stream Privacy for Grassroots Participatory Sensing

Raghu K. Ganti, Nam Pham, Yu-En Tsai, and Tarek F. Abdelzaher
Department of Computer Science, University of Illinois, Urbana-Champaign
rganti2,nampham2,ytsai20,zaher@illinois.edu

By:Siddharth Mohan
EEL 6788

Outline

- Assumption
- Poolview
- Data Perturbation
- Algorithms
- Attacks & Measures
- Context privacy
- Traffic Analyzer
- Participatory sensing +Future work

What is this paper all about

- In order to provide privacy guarantees on stream data in grassroots participatory sensing applications various mathematical foundations & architectural components are developed.
- The structure of these components or applications makes it hard to enforce privacy

Assumption

- In this application we consider communities of individuals with sensors collecting streams of private data for personal reasons. These data could be of value if shared with the community for fusion purposes to compute aggregate metrics of mutual interest.

- Main problem in such applications-Privacy???

- Solution-They adopted a client server architecture,called pool view, where clients share(perturbed) private sensory data and servers (called pools aggregate such data into useful information made available to the community

Poolview

- A privacy preserving architecture based on data perturbation on the client-side to ensure individuals'privacy.
- Poolview uses community-wide reconstruction techniques to compute the aggregate information of interest.
- Grassroots applications:They refer to those initiated by members of the community themselves as opposed to by some governing or official entities.

Data Perturbation

- In a participatory sensing application collected data is shared (perturbed)form to compute community statistics.
- Mathematical foundations needed for perturbing time-series data in grassroots participatory sensing applications are initially provided.
- The two basic conditions to be considered while perturbation of a user's sequence are

1.)The individual data items and their trend (i.e., their changes with time) cannot be estimated without large error.

2.)The distribution of community data at any point in time, as well as the average community data trend can be estimated with high accuracy.

Data Perturbation Algorithm

- Generally the user data streams can be generated in two ways

Linear discrete model

Non Linear discrete model

- An example of diet tracker application is taken into consideration

- Parameters involved:λ_k,β,W_0
- β-Body Metabolsism coefficient

- W_0-Initial weight of the person before dieting

- λ_k-Average calorie intake of the person on day k

Algorithm

- The weight W_k of a dieting user on day k of the diet can be given by a non linear equation
- $W(k) = W(k-1) + \lambda_k + \beta W(k-1)^{3/4}$
- $W(0) = W0$
- The model parameter vector is $\boldsymbol{\theta = (\beta, W_{0)}}$
- From the above given equations one can easily determine the accurate weight of a user.

Algorithm

- The model for a dieting person is not private and the probability distributions of weight parameters over a large population can be approximately hypothesized.
- It is desired to hide the parameters θ and u(input) of any given user from being estimated. This protects an individual user's privacy.

Privacy & user data reconstruction

- Privacy breaches are different depending on the exploitation method employed by the adversary.
- Quantify (privacy) by analyzing the degree to which actual user data can be estimated from perturbed data using methods that take advantage of data correlations such as PCA and spectral filtering.
- Estimation methods such as Maximum Mean Squared Estimation (MMSE) can also be involved

Attacks & Countermeasures

Attacks	Countermeasures
Reconstruction attacks using estimation techniques(ex, PCA)	Use a noise model similar to data model
Malicious noise models from server	Rejects models that do not fit data on client side
Sequential noise models from server	Limit noise model updates accepted by client
Malicious data from client	Reject outliers by servers

Context privacy

- Data measurements are associated with a given time & place
- For example, data measurements are associated with a given time and place. Sharing data (e.g., on city traffic), even in perturbed form still puts the user at a given time at a particular location.
- Realistic example- A user may reveal that he/she was on a particular city street at 11am on a Wednesday but not reveal which Wednesday it was. This could be enough to achieve a level of privacy and at the same time satisfy the statistical need of the aggregation server, say, if the statistic being computed is that of traffic density as a function of time of day and day of week.

Traffic Analyzer

- Analysis of patterns such as rush hour traffic, off-peak traffic, average delays between different points in the city as a function of time of day and day of the week.
- The average speeding statistics on selected streets can shed light on traffic safety and traffic congestion status both at a given point in time and historically over a large time interval.

Traffic Analyzer

- This paper evaluated a case study –Traffic Analyzer
- Traffic Analyzer can be summarised through the following steps-

- The individual user connects to the information distillation server

-The server sends an HTTP POST request to the user's personal storage server asking for the requisite data.

- The request is intercepted by the user's privacy firewall. The request is validated by the user's privacy firewall by first authenticating it to ascertain if it is from the correct server and then if that server has valid access rights to the data that is requested.

- Data are then shared in a perturbed manner

Traffic Analyzer

- Data collected from a community of 30 users with usage of GPS device(Garmin)
- Sampling frequency in this study was 1 sample every 15 seconds
- In a stretch of 1.3 miles data was gathered at noon & evenings
- The average speed on each of the two considered roads is calculated from perturbed user data.
- Perturbation scheme employment requires a noise model which is given as $f_k = A_o + Ea_i \sin(b_i * k + c_i)$

Coping with malicious servers & users

- A malicious server is one that “cheats” by announcing a poor noise model in an attempt to get poorly perturbed user data such that user privacy can be violated.
- The range of the malicious data has no effect on the reconstruction error.

Future work

- An architecture for participatory sensing, called Partisans, has been proposed over the years in which main challenges addressed are those of data verifiability and privacy.
- Architecture to adress privacy issues when multi modal data are shared
- More emphasis on context privacy

Conclusion

- Future of sensing systems
- Applications including a participatory sensor network to search & rescue hikers in mountains,vehicular sensor networks(CarTel), BikeNet have already been published
- These applications lack in a number of factors:

i)Concern over privacy

ii)Reliability

- Past work in this field involved random perturbration,randomized response,secure multi party computation.

Difficulties

- The paper is vague in certain parts & too many mathematical derivatives make it harder
- The real time examples of Diet tracker & Traffic Analyzer were explained with accuracy & was understandable
- Reconstruction accuracy was unclear.

Questions