Personal Data Vault: a privacy architecture for mobile personal sensing

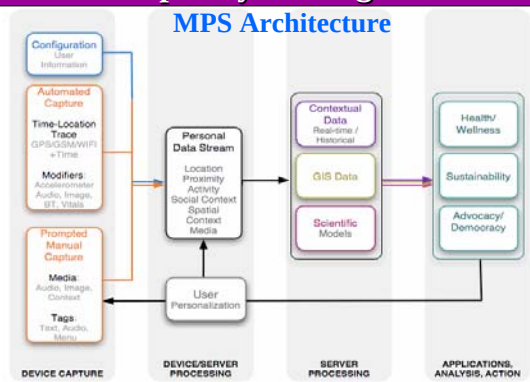
Min Mun, Katie Shilton, Kenny Guan, Gene Auyeung
Nicolai Peterson, Jeff Burke, Deborah Estrin, Mark Hansen, Jerry Kang
Center for Embedded Networked Sensing – http://urban.cens.ucla.edu

Introduction: Mobile Personal Sensing, a Platform for Participatory Sensing

Mobile Personal Sensing

Mobile Personal Sensing (MPS) is a platform for participatory sensing with which people use ***mobile phones*** to ***record*** and ***transmit*** sound, images, location, motion data, and ***web services*** to ***aggregate*** and ***Interpret*** the assembled information

MPS Architecture



Problem Description: Continuously Recording Location and Activity is Potentially Invasive

The Data Gathered through MPS is ***personal***, as well as being ***valuable***; it quantifies habits, routines, associations, and is easy to mine

An Architecture to Support
- ***Protecting Individual privacy***
- ***Documenting ownership***
- ***Providing visibility of processing***

is Required.

Proposed Solution: Personal Data Vault, A Privacy Architecture for MPS

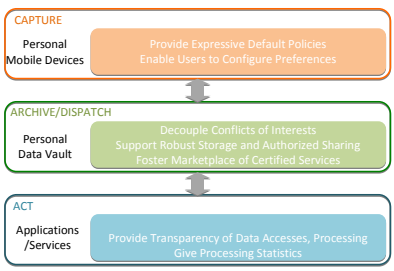
Application-centric Architecture

- Loss of data ownership, data lock-in
- Inefficient, hard to scale
- Data redundancy
- Conflict of interest

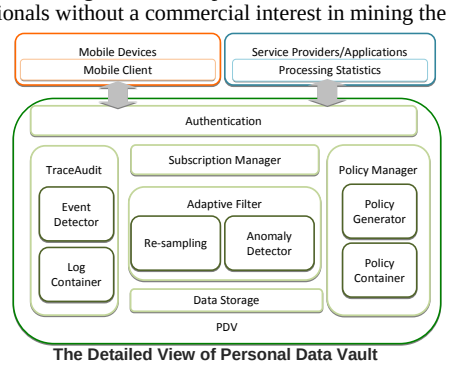
Mobile-centric Architecture

- Limitation of resources
- No data backup

Personal Data Vault Model



- Enables individuals to control and manage their data
- Allows users to make and revoke decision to share filtered or derived data
- 'Professionalizing' data vault operations: Intended to be run by professionals without a commercial interest in mining the data



The Detailed View of Personal Data Vault

Basic Services

- Authentication to control accesses
- Authorization to manage data feeds to service providers
- View, delete and update to deal with data

TraceAudit

- To foster a market place of trustworthy service providers and provide transparency of processing
- All applications should inform the PDV of any activities on the data obtained from PDV
- Suspicious events can be detected by event detector

Adaptive Filter

- Filtering policies adapt to input from users and service providers
- Three methods: Error-tolerant data sampling, mobility anomaly detection, data interpolation
- **Error-tolerant Data Sampling**:
  - The high degree of data corruption required to preserve privacy could preclude the altering of application outputs as a means of supporting privacy aware location services.
  - Identifies the sampling method that minimize the amount of information shared, while minimizing the perturbation of inferences made by service providers.
  - Feedback or processing statistics from users and service providers play an important role.

An Example of Processing Statistics - Precision and Recall Confusion Matrices of The Activity Classifier using Various Sensors

	Being Stationary		Walking		Driving		All	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Precision
GSM, Wi-Fi	66.3%	92%	84.2%	66.1%	90%	82.9%	80.2%	80.3%
GSM	76.3%	70.7%	59.4%	71.2%	80.3%	68.8%	72%	70.23%
Wi-Fi	75.9%	60.6%	64.6%	61.4%	70.1%	84.3%	70.20%	68.77%
GPS	81.6%	92.5%	93.5%	91.2%	98.9%	91.3%	91.3%	91.67%

- **Mobility Anomaly Detection:** Abnormal activities such as unusual route can be detected and reported to service providers with users' permission
- **Data Interpolation**: It determines positions in-between the samples and quantifies the position uncertainty.

Sampling Error using Linear Interpolation

$$\mathcal{P}_2(x,y) = \begin{cases} \frac{1}{A} & \text{for } x^2+y^2 \le r_1^2 \wedge (x-s)^2+y^2 \le r_2^2 \\ 0 & \text{otherwise} \end{cases}$$