

Privacy and Security : Research

—

Wendy Vasquez

Overview

Privacy does not mean anonymous

Methods for Data Privacy

Hardware and Embedded Privacy Methods

Do you trust your devices?



Ways to acquire your location via mobile devices (the cell-phone):

- GPS
- Wi-Fi
- Cellular Network

Did you give consent ?
Is it really anonymous data ?

Location Based Services

What's App

text messages, images, location etc..

App Develops

2012 suit against Apple, Twitter, Facebook , Instagram for distributing privacy -invading mobile applications

Android, Apple and Window Phones

TaintDroid project

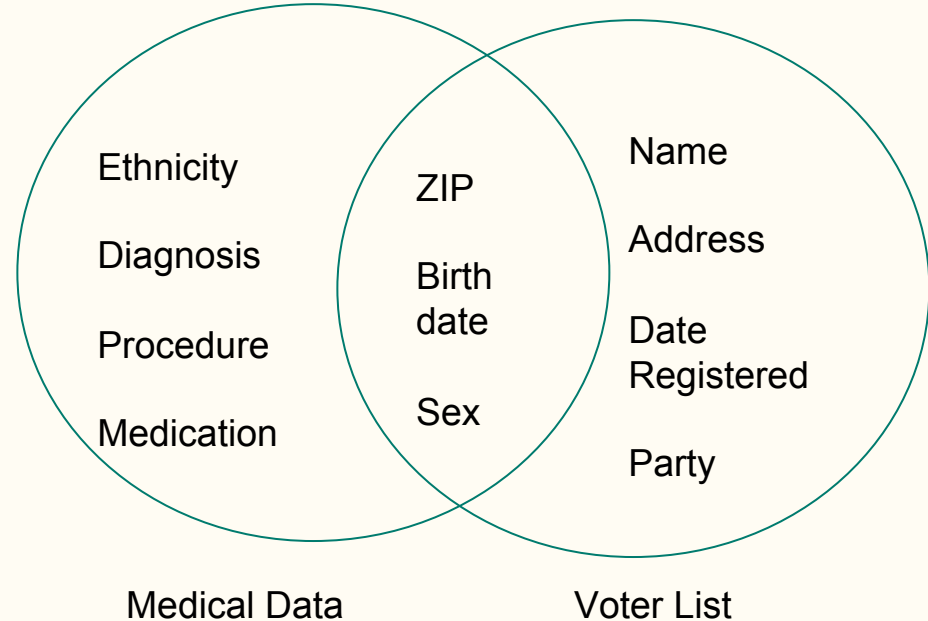
15 application sending users gps information to third party

7 of 30 send hardware unique identifier SIM serial number

Privacy does not mean Anonymous

Linking or matching data

“anonymous” data is used to re-identify individuals by linking or matching the data to other data or by looking at unique characteristics found in the release data .



Wouldn't it be nice ...

a formal framework :

construction and evaluation algorithm and systems that release information

to establish limits what can be of the entities that are to be protected.



Some Methods for Data Privacy

Suppression

Generalization

Query Restriction

Suppression

removes information that is
sensitive or can be used to
infer the sensitive information
are not released

Name	Gender	Age
*	m	14
*	f	25

Generalization

replace specific information
with ranges

i.e. generalize :)

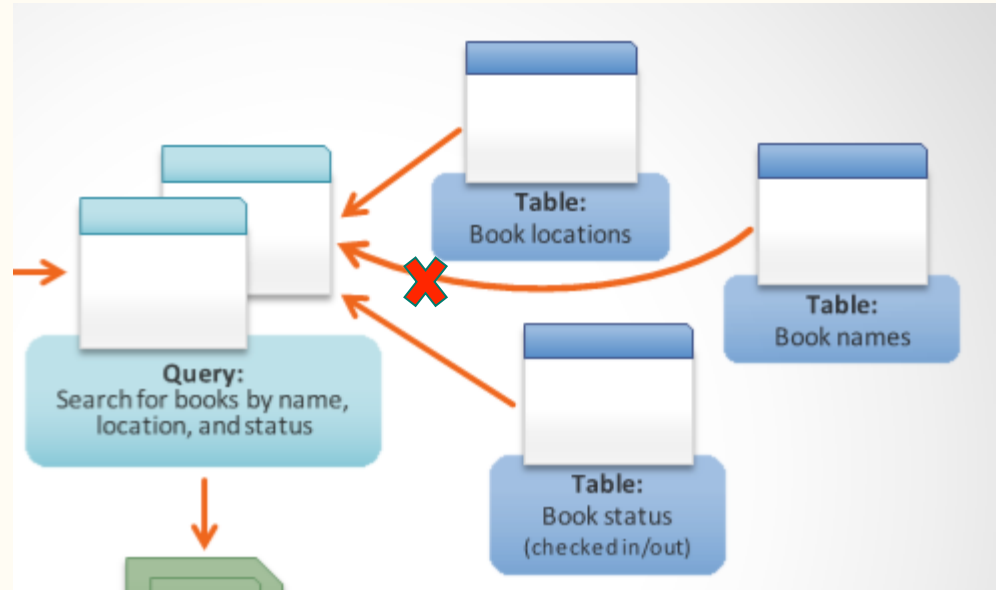
Name	Gender	Age
*	m	14
*	f	26



Name	Gender	Age
*	m	< 20
*	f	> 25

Query Restriction

prohibits queries that can reveal sensitive information by suppressing inference to the sensitive information



A formal framework

K- anonymity protection model

Definitions - k- anonymity protection model

Let $RT(A_1, \dots, A_n)$ be a table and QI_{RT} be the quasi-identifier associated with it. RT is said to do satisfy k-anonymity if and only if each sequence of values in $RT[QI_{RT}]$ appears with at least k- occurrences in $RT[QI_{RT}]$.

attributes are suppressed or generalized until each row is identical with at least k-1 other rows. At this point the database is said to be k-anonymous.

What is a quasi-identifier

pieces of information that are not of themselves unique identifiers, but are sufficiently well correlated with an entity that they can be combined with other quasi-identifiers to create a unique identifier

Gender
Zip Code
Age

K- anonymity example : $k = 2$

Name	Birth	Gender	Problem
A	1993	m	chest pain
B	2008	f	hypertension



Birth	Gender	Problem
1993	m	chest pain
2008	f	hypertension



k- anonymity

Birth	Gender	Problem
1993	m	chest pain
2008	f	hypertension
1993	m	chest pain
2008	f	hypertension

Attacks on k-anonymity

Unsorted matching - identifying based on position in table

Complementary release - when quasi-identifier is a subset of attributes

linking on non - QI information result on the result information

Temporal attack - when data changes , different RT_t can be join to reveal sensitive information

A method of a specific solution

Personal- data linking or matching data !

Hardware and Embedded Security

—

Privacy

Attack :

eavesdropping communication between devices communication and retrieve private information

Defense:

all levels of communication between devices have a lightweight encryption

Trust Management

Can the device that is transmitting the data be trusted?

Attack:

Attacker attempts to exploit the vulnerabilities of an application on a node to corrupt data or manipulate actuators

car manipulation

Defense:

Leverage hardware performance counters and PUFs to detect tampering

Identity Management

Is a sensor authorized to send data to the node?

Attack:

Device A fakes the identity of Device B

Defense:

Use PUF and different challenge-response pairs to find true identification