Who owns Your Data

Prashanth Mohan
Privacy Threats

Intentional or unintentional misuse

AOL search log, Netflix contest

Insider attacks

Software vulnerabilities

Google Fires Employee for Snooping on Users

The Internet search giant says the software engineer broke its "strict internal privacy policies." He allegedly accessed information about four teenagers.

One Of The 32 Million With A RockYou Account? You May Want To Change All Your Passwords. Like Now.
What is the “Smart” Grid?

- Residential Locations
- Data Center
- Industrial Factories
- Generation Facilities
What are we not talking about?

- Data integrity associated with collection
- Data transfer protocols
- Security of devices
The proposed rules would require utilities to provide third parties with access to usage data [...] The PUC rejected suggestions that third parties should be required to register for certification to offer services that require access to customer energy consumption data. -- CPUC on Smart Grid Privacy and Security (May 9th, 2011)

Once energy consumption information flows outside of the home, the following questions may come to the minds of consumers: Who will have access to this intimate data, and for what purposes? Will I be notified? What are the obligations of companies making smart appliances and Smart Grid systems to build in privacy? How will I be able to control the details of my daily life in the future? -- Office of the Information and Privacy Commissioner, Ontario
Data mining and Intelligence

Silo based data services

[Diagram showing data departments and connections]

- Confidential
- PG&E

[Other department icons and connections]
The internet isn’t very different!
What is the privacy risk?

Sensor Database

Record 1

Records ...

Record n

Std Dev, k-means, average, k-NN, etc
What is the privacy risk?

Sensor Database

Record 1

Records ...

Record n

I want information about Batman’s whereabouts

Objective:
Do not leak “too-much” information about Batman’s location
What is the privacy risk?

From the output of the privacy mechanism, DB1 and DB2 should be indistinguishable.
Hiding in the Crowd - GUPT

Open Source – BSD Licensed
http://github.com/prashmohan/GUPT

Abhradeep
Elaine
Dawn
David
Differentially Private operation
Differential Privacy

\[ \Pr \left[ f(D) \in S \right] \leq \exp(\varepsilon) \times \Pr \left[ f(D') \in S \right] \]

Computation

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

Function Sensitivity

\[ f(x) = f(x) + \text{Lap} \left( \frac{\Delta f}{\varepsilon} \right) \]

Privacy Allowance
1. Computation
2. Accuracy

Differentially
Private Answer

Web Frontend

1. Data Set
2. Privacy Budget ($\epsilon$)

Data Analyst

Data Owner

Comp Mgr XML RPC Layer

Untrusted Computation

Isolated Execution Chambers

Isolated Execution Chambers

Isolated Execution Chambers

Data Set Manager

Computation Manager
Normalized Intra Cluster Variance

- GUPT-tight
- GUPT-loose ($\gamma=1$)
- GUPT-loose ($\gamma=6$)
- Baseline ICV

Privacy Budget ($\epsilon$)
Trust in today’s cloud

Platform provider

Minimally Trusted

Trusted

Application provider

Untrusted

Trusted

Untrusted
Platform for Private Data

Petros

Babis

Elaine

Dawn

Emil

Mohit
<table>
<thead>
<tr>
<th>Code Base</th>
<th>Trust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware</td>
<td>Trusted</td>
</tr>
<tr>
<td>tboot</td>
<td>Trusted</td>
</tr>
<tr>
<td>BIOS + Bootloader</td>
<td>Untrusted</td>
</tr>
<tr>
<td>Storage Provider</td>
<td>Untrusted</td>
</tr>
<tr>
<td>PPD Platform</td>
<td>Trusted</td>
</tr>
<tr>
<td>Linux Kernel</td>
<td>Trusted (in Prototype)</td>
</tr>
</tbody>
</table>
Feedback

• Does Privacy and Security concerns limit your business processes and in what ways?
• If and how does your institution define privacy risks/issues?
Thank you!
Backup Slides
1. Launch app
2. Launch container for Alice
3. Intent to edit file
4. Request trusted UI
5. Choose file to edit
6. Mount file in container
7a. Edit file
7b. Read and write to file

ACLS
<table>
<thead>
<tr>
<th>id</th>
<th>o</th>
<th>r</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Integrity check
<table>
<thead>
<tr>
<th>id</th>
<th>Hash</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>@#%</td>
</tr>
</tbody>
</table>

TCB: PPD Controller

Data capsules

Untrusted Storage
Simple data anonymization?
System throughput [req/sec]

Number of users

Friendshare on Linux, SR
Friendshare on PPD, SR
<table>
<thead>
<tr>
<th>Feature</th>
<th>GUPT</th>
<th>PINQ</th>
<th>Airavat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Works with unmodified programs</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Allows expressive programs</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Automated privacy budget allocation</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Protection against privacy budget attack</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Protection against state attack</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Protection against timing attack</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Resampling factor ($\gamma$)

Normalized Intra Cluster Variance

GUPT-loose

Baseline ICV
Accuracy

GUPT-tight ($\gamma=1$)
Non private baseline

Privacy Budget ($\epsilon$)
Resampling Factor (γ)

Time (seconds)

- Non Private baseline
- GUPT-helper
- GUPT-loose
Normalized privacy budget lifetime

- **GUPT-helper constant $\epsilon=1$**
- **GUPT-helper variable $\epsilon$**
- **GUPT-helper constant $\epsilon=0.3$**