Protecting Consumer Privacy in Smart Metering by Randomized Response

Bastian Stölb and Josep Domingo-Ferrer
Universitat Rovira i Virgili

bastian.stoelb@estudients.urv.cat
josep.domingo@urv.cat
Smart Meter spreading in Europe

2018: 44%  
2023: 71%

source: www.berginsight.com
Serious privacy concerns due to Smart Meters

**PRO**
- real-time monitoring
- matching production + consumption

**CONTRA**
- number of people in an accommodation
- invitation for burglary
- NALM: observing detailed consumption
Privacy models

Randomized Response (1965)

k-Anonymity (1998)

Differential Privacy (2006)

Permutation Paradigm (2016)
Data Aggregation: reducing costs for computation + communication

Erkin + Tsudik in [10]
- Spatial aggregation: (geographically) clustered. Individual households are protected. Load-balancing is possible.
- Temporal aggregation: smart meters withhold data until a specific time interval has passed.
- Spatio-temporal aggregation: hybrid setting.
G. Ács + C. Castelluccia: I Have a DREAM! (DiffeRentially privatE smArt Metering)

- Straightforward, economic + easy to implement.
- No Trusted Third Party.
- Laplacian noise is added + stream cipher is applied. ➔ Quite low computational costs.
- SMs clustered + aggregated.
- Disadvantages:
  - coarse resolution,
  - does not cope with malfunctions.
01 Introduction
02 Non-Crypto Privacy
03 Related Work
04 Proposed Scheme
05 Results
06 Conclusion
individual readings with (homomorphic) aggregation

time-aggregated readings (noise is added to them)

Non-Crypto

collect + protect non-aggregated individual readings (with RR?)
Randomized Response: plausible
deniability – a closer look

- $X$ attribute containing the answer to sensitive question.
- If $X$ can take $r$ possible values, then $Y$ (RR value) follows an $r \times r$ matrix of probabilities

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & \ddots & \vdots \\ p_{r1} & \cdots & p_{rr} \end{pmatrix}$$

where $p_{uv} = Pr(Y = v|X = u)$, for $u, v \in 1, \ldots, r$ denotes the probability that RR is $v$, when it is $u$. 
Randomized Response: plausible deniability – a closer look (2)

- Let $\pi_1, \ldots, \pi_r$ be the proportions of respondents whose true values fall in each of the $r$ categories of $X$ and let $\lambda_v = \sum_{u=1}^{r} p_{uv} \pi_u$ for $v = 1, \ldots, r$, be the probability of the reported value $Y$ being $v$.

- If we define $\lambda = (\lambda_1, \ldots, \lambda_r)^T$ and $\pi = (\pi_1, \ldots, \pi_r)^T$, it holds that $\lambda = P^T \pi$.

- If $\hat{\lambda}$ is the vector of sample proportions corresponding to $\lambda$ and $P$ is nonsingular, in [5] it is proven that an unbiased estimator $\hat{\pi}$ can be computed as $\hat{\pi} = (P^T)^{-1} \hat{\lambda}$. 
Creating the RR matrix

- RR probability matrices with $p = 0.4$, $p = 0.6$, $p = 0.8$ in the main diagonal
- 3 different attenuation formulae A, B and C on $p$
- Rescale the probabilities in each row of the matrix for them to sum to 1!

\[
P_A = \begin{pmatrix}
p & p/2 & p/4 & p/8 \\
p/2 & p & p/2 & p/4 \\
p/4 & p/2 & p & p/2 \\
p/8 & p/4 & p/2 & p \end{pmatrix}
\]

\[
P_B = \begin{pmatrix}
p & p/2 & p/3 & p/4 \\
p/2 & p & p/2 & p/3 \\
p/3 & p/2 & p & p/2 \\
p/4 & p/3 & p/2 & p \end{pmatrix}
\]

\[
P_C = \begin{pmatrix}
p & p^2 & p^3 & p^4 \\
p^2 & p & p^2 & p^3 \\
p^3 & p^2 & p & p^2 \\
p^4 & p^3 & p^2 & p \end{pmatrix}
\]
Analysis

Graph A: Graphs for different values of $p$.
- Blue: $\pi$
- Orange: $\hat{\pi}$ for $p = 0.4$
- Gray: $\hat{\pi}$ for $p = 0.6$
- Yellow: $\hat{\pi}$ for $p = 0.8$

Graph B: Graphs for different values of $p$.

Graph C: Graphs for different values of $p$.
In each diagram the blue line shows the frequencies of $\pi$, the orange line the frequencies of $\hat{\pi}$ for $p \approx 0.4$, the gray line $\hat{\pi}$ for $p \approx 0.6$ and the yellow line $\hat{\pi}$ for $p \approx 0.8$.

The values of $p$ are approximate due to rescaling.

Lower values of $p$ $\Rightarrow$ more privacy, less accuracy

$p = 0.6$ offers relatively high privacy, while maintaining accuracy at the same time.

For attenuation formulae A and B the difference of the three values of $p$ is small.

For attenuation C, accuracy clearly increases with $p$. 

<table>
<thead>
<tr>
<th>Agenda</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>01</strong> Introduction</td>
</tr>
<tr>
<td><strong>02</strong> Non-Crypto Privacy</td>
</tr>
<tr>
<td><strong>03</strong> Related Work</td>
</tr>
<tr>
<td><strong>04</strong> Proposed Scheme</td>
</tr>
<tr>
<td><strong>05</strong> Results</td>
</tr>
<tr>
<td><strong>06</strong> Conclusion</td>
</tr>
</tbody>
</table>
Conclusion

- Trade-off between privacy + accuracy.
- We will quantify the privacy achieved in terms of differential privacy.
- Given a range for the SM readings, we will explore the best number of intervals to split that range, trade-off between accuracy + computational complexity.
PROTECTING DATA
RANDOMIZED RESPONSE
INDIVIDUAL READINGS
LOW OVERHEAD
WITHOUT AGGREGATION
NON-CRYPTOGRAPHY
ACCURACY