Utility-Privacy Trade-Offs of Data Manipulation Techniques for Smart Metering

Liang Cheng, Ph.D.

Department of Computer Science and Engineering
Lehigh University
Bethlehem, Pennsylvania 18015

cheng@lehigh.edu
Smart meters collecting, processing, storing, and reporting users’ energy consumption data with high fidelity

Allow utility customers to easily and securely access their usage information in a **consumer-friendly** and **computer-friendly** format and control data disclosure.
Energy Disaggregation Using Green Button Data

- **Markov Model algorithms**
  - Factorial Hidden Markov Models
  - Conditional Factorial Hidden Markov Models
  - Conditional Factorial Hidden Semi-Markov Models

- **K-Nearest Neighbor (KNN) algorithm**

- **Support Vector Machine (SVM) algorithm**

Algorithm Comparisons

- **SMART METER**
- **GREEN BUTTON**
- **ENERGY DISAGGREGATION**
- **PRIVACY PROTECTION**
- **CONCLUSION**

---

EnergyPath 2019, July 25
Prof. Liang Cheng: [http://liangcheng.info](http://liangcheng.info)
Algorithm Comparisons

- Precision
  - TP/(TP+FP)
- Recall
  - TP/(TP+FN)

\[ F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \]
Non-intrusive Load Monitoring

Real-Time Itemized Electricity Consumption Intelligence for Military Bases by Omid Jahromi and Alan Meier, NILM Workshop 2018

- **Recommendations:** Install CO2 sensor to control ventilation (estimated saving of 40% ventilation), Install LED lighting & motion sensors (estimated saving of 20% lighting), power-manage office equipment (e.g. disable screensavers, estimated saving of 0-20% office equipment)


Load Disaggregation of Industrial Machinery Power Consumption Monitoring Using Factorial Hidden Markov Models by Pedro Martins, Pedro Bittencourt, and Raphael Pinto, NILM Workshop 2018

* NILM Workshop: http://nilmworkshop.org
** EU NILM Workshop: http://www.nilm.eu

and many more …
NILM + Anomaly Detection

Clustering

- Grouping based on day-of-week information
- Clustering algorithm
- days labeled by their clusters

Green Button data (coarse-grained)

Outlier Detection

- Grouping into clusters
- Feature selection
- Outlier detection algorithm

Green Button data (fine-grained)

Anomaly detection

- Green Button data of time duration containing anomalies
- NILM algorithm
- Estimated appliance-specific energy consumption during occurrences of anomalies

NILM algorithm

- Estimated appliance-specific energy consumption of suspicious periods of time
- Anomaly detection
- Specific time intervals related to abnormal energy consumption
Sensitive information can be extracted from appliance-specific energy usages.

- **Occupancy states**

- **User activity patterns**

- **Multimedia contents being played on a TV set**
Privacy Protection Techniques

- **Encryption-based** techniques

- **Battery-based load hiding (BLH)** techniques

- **Data manipulation** techniques
How well can data manipulation techniques prevent leakage of appliance-level energy consumption information?

When are investments on BLH techniques necessary to protect privacy?

Adversary model

Data utility model

Privacy model

Definition 1 (Data Utility Metric): Given two time series $X_i^T$ and $\hat{X}_i^T$ for appliance $i$, the distortion between $X_i^T$ and $\hat{X}_i^T$ can be measured by their distance $d(X_i^T, \hat{X}_i^T)$. Suppose that there are $N$ samples in $X_i^T$ (and $\hat{X}_i^T$), we use the average distortion $\bar{d} = \frac{d(X_i^T, \hat{X}_i^T)}{N}$ as the utility metric for $i$.

Definition 2 (Privacy Metric): Given two time series $X_i^T$ and $\hat{X}_i^T$ over the same time period $T$ for appliance $i$, the mutual information $I(X_i^T, \hat{X}_i^T)$ between the two series is

$$I(X_i^T, \hat{X}_i^T) = \sum_{x \in X_i^T} \sum_{y \in \hat{X}_i^T} \ln \frac{p(x,y)}{p(x)p(y)},$$

where $p(x)$ and $p(y)$ are the probability density functions of random variables $x \in X_i^T$ and $y \in \hat{X}_i^T$, and $p(x,y)$ is the joint probability density function.

The greater the distortion, the less the utility.

The greater the mutual information, the more the privacy leakage.
Experiment Settings

- 1 house, 13 appliances
- Sampling rates (<=1/3 Hz)
- Training set (1 week) and Testing set (11 days)
- Random noise w/ uniform distribution
- FHMM algorithm
- A 50W bin size for computing mutual information

<table>
<thead>
<tr>
<th>Index</th>
<th>Appliance</th>
<th>Index</th>
<th>Appliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>oven</td>
<td>2</td>
<td>refrigerator</td>
</tr>
<tr>
<td>3</td>
<td>dishwasher</td>
<td>4</td>
<td>kitchen-outlets-1</td>
</tr>
<tr>
<td>5</td>
<td>lighting-1</td>
<td>6</td>
<td>washer-dryer</td>
</tr>
<tr>
<td>7</td>
<td>microwave</td>
<td>8</td>
<td>bathroom-gfi</td>
</tr>
<tr>
<td>9</td>
<td>electric-heat</td>
<td>10</td>
<td>stove</td>
</tr>
<tr>
<td>11</td>
<td>kitchen-outlets-2</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>lighting-3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conclusion

• NILM / Energy disaggregation may enhance energy consumption awareness and enable additional smart grid applications

• NILM / Energy disaggregation may reveal private information of energy consumers

• Users can balance between privacy revelation and data utility by choosing the proper data manipulation techniques
  • Although extra investments on batteries and control infrastructure are not required, the granularity of control supported by these techniques is coarse

This work was supported by PPL Corporation and Lehigh University. Any opinions, findings, and conclusions or recommendations expressed in this talk/paper are those of the author(s) and do not necessarily reflect the views of the sponsors of the research.