sensorSift
Balancing Utility and Privacy in Sensor Data
Rise of \{Sensors + AI\}

- People expect rich computational experiences to be available in every context

As a result, our world is increasingly visible to intelligent computers
  - Minimal cost of sensors
  - Cheap computational power
  - Advances in machine reasoning
Lack of Balance

- There are many **benefits** of smart-sensor applications
  - *Increased Productivity, Connectivity, and Interactivity*
- However there are also potential **negative** effects
  - *Privacy Risks*
Goals

• Develop a quantitative framework for **balancing** privacy and utility in smart sensing applications.
  – Empower users with privacy guarantees
  – Applications retain functionality

• Evaluate the quality of our framework against state of the art machine inference

• Offer a flexible solution so that the future demands of users/applications can be supported
Usage Model 1

Sensor data releases to smart applications are often **risk carrying**.

**Common Practice:** Sensor releases all of the raw data to an Application (e.g. MS Kinect)

Sensor :{ 1 sensor data } \(\rightarrow\) App :{ 2 feature extract, 3 classify, 4 logic}
Usage Model 1

Sensor data releases to smart applications are often **risk carrying**

**Common Practice:** Sensor releases all of the raw data to an Application (e.g. MS Kinect)

Sensor :{ 1 sensor data } → App :{ 2 feature extract, 3 classify, 4 logic}
Usage Model 2

Sensor data releases to smart applications are often **arbitrarily stifling**

**Common Practice:** Only a predefined set of features is available to an Application (e.g., iOS)

**Platform** :\{ 1 sensor data , 2 feature extract, 3 classify \}  \rightarrow  **App** :\{ 4 logic \}
Usage Model 2

Sensor data releases to smart applications are often **arbitrarily stifling**

**Common Practice:** Only a predefined set of features is available to an Application (e.g., iOS)

**Platform:** `{1 sensor data, 2 feature extract, 3 classify} → App :{4 logic}`

- **INNOVATION**
- **++ PRIVACY**
Solution

• Users choose what attributes to keep **private**
• Applications can request non-private (**public**) attributes
  – Public attributes can be invented!
Solution

- Users choose what attributes to keep **private**
- Applications can request non-private (**public**) attributes
  - Public attributes can be invented!
- We transform (sift) sensor data to reveal the **public** but hide the **private** attributes

**Plat.** : {1 sensor data, 2 sift features} → **App** {3 classify, 4 logic}

+ INNOVATION
+ PRIVACY
Evaluation Context

ATTRIBUTES: visually describable characteristics about a face
System Overview

Scenario:

- **USER**: I don’t want apps. to have knowledge about my **race** and **gender**
- **APPLICATION**: Is the user **smiling**?

  > **POLICY**: PRIVATE {race, gender}, PUBLIC {smiling}

System:

1. Generates Sift
2. Verifies Sift
3. Applies Verified Sift
**System Overview**

**Scenario:**
- **USER:** I don’t want apps. to have knowledge about my race and gender
- **APPLICATION:** Is the user smiling?

> **POLICY:** PRIVATE {race, gender}, PUBLIC {smiling}

**System:**
1. Generates Sift
2. Verifies Sift
3. Applies Verified Sift
Generating Sifts

Intuitively, sifting finds the safe region(s) in feature space which are in the public feature set $B$ but not in the private one $A$.

Feature regions are based on a large database of sensor samples.

$A = $ eyewear (private)
$B = $ gender (public)

SAFE
OVERLAP
(UNSAFE)

gender
eyewear
safe region
Generating Sifts

Intuitively, sifting finds the safe region(s) in feature space which are in the public feature set \( B \) but not in the private one \( A \).

\[ A = \text{eyewear (private)} \]
\[ B = \text{gender (public)} \]

Safe region(s) may not always exist for certain attribute correlations.
Sifting Details

\[ X = \text{Raw Features} \]
\[ X' = \text{Sifted Features} \]

\[ X_n, n > 100k \]

\[ X'_n, n \sim 5 \]

PPLS

**Algorithm 1: Privacy Partial Least Squares**

1. Set \( j = 0 \) and cross-product \( S_j = X^T Y^+ \)
2. If \( j > 0 \), \( S_j = S_{j-1} - P(P^T P)^{-1} P^T S_{j-1} \)
3. Compute the largest eigenvector \( w_j \):
   \[
   \left[ S_j^T S_j - X^T Y^- (Y^-)^T X \right] w_j = \lambda w_j
   \]
4. Compute \( p_j = \frac{X^T X w_j}{w_j^T X^T X w_j} \)
5. If \( j = k \), stop; otherwise let \( P = [p_0, \ldots, p_j] \) and \( j = j + 1 \) and go back to step 2

\[ \text{find} \quad \max_w \left[ \text{cov}(Xw, Y^+)^2 - \lambda \ast \text{cov}(Xw, Y^-)^2 \right] \]

\( Y^+ = \text{labels of public attribute(s)} \)
\( Y^- = \text{labels of private attribute(s)} \)
Performance Metrics

- A successful sift will have low scores on both **PubLoss** and **PrivLoss**

  - **PubLoss**: Decrease in sifted public attribute classification accuracy relative to the achievable accuracy using raw (unsifted) data.

  - **PrivLoss**: Gain in sifted private attribute classification accuracy relative to chance.

\[
\text{PubLoss} = ML_m(X, Y^+) - ML_m(\text{PM}_{Y^+(X,K)}, Y^+)
\]

\[
\text{PrivLoss} = ML_m(\text{PM}_{Y^+(X,K)}, Y^-) - .5
\]

*Classifiers: Linear Support Vector Machine (SVM), Non-Linear SVM, Neural Network, Random Forest, kNearest Neighbors*
Dataset & Attributes

**PubFig** Database ~45,000 face images of 200 celebrities, 72 attributes

**Attributes** are [binary] labels for visually describable characteristics,

**Attribute Clusters**
- Wavy Hair
- Arched Eyebrows
- Wearing Lipstick
- Blond Hair
- Youth

Male - **M**, Attractive Female - **AF**, White - **W**,
Youth - **Y**, Smiling - **S**, Frowning - **F**, No Eyewear - **nE**,
Obstructed Forehead - **OF**, No Beard - **nB**, and Outdoors - **O**.
<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>AF</th>
<th>W</th>
<th>Y</th>
<th>S</th>
<th>F</th>
<th>nE</th>
<th>OF</th>
<th>nB</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0</td>
<td>32</td>
<td>13</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>17</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>AF</td>
<td>21</td>
<td>0</td>
<td>2</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>14</td>
<td>4</td>
</tr>
<tr>
<td>W</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Y</td>
<td>17</td>
<td>22</td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>S</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>10</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>F</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>nE</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>OF</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>nB</td>
<td>2</td>
<td>9</td>
<td>4</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

**PubLoss**

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>AF</th>
<th>W</th>
<th>Y</th>
<th>S</th>
<th>F</th>
<th>nE</th>
<th>OF</th>
<th>nB</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>9</td>
<td>8</td>
<td>4</td>
<td>11</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>AF</td>
<td>9</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>W</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Y</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>S</td>
<td>8</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>nE</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>OF</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>nB</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>O</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

**PrivLoss**

**Correlation**

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>AF</th>
<th>W</th>
<th>Y</th>
<th>S</th>
<th>F</th>
<th>nE</th>
<th>OF</th>
<th>nB</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1</td>
<td>-0.68</td>
<td>-0.05</td>
<td>-0.36</td>
<td>-0.19</td>
<td>0.16</td>
<td>-0.19</td>
<td>0.02</td>
<td>-0.72</td>
<td>0.05</td>
</tr>
<tr>
<td>AF</td>
<td>-0.68</td>
<td>1</td>
<td>0.15</td>
<td>0.5</td>
<td>0.19</td>
<td>-0.14</td>
<td>0.33</td>
<td>-0.14</td>
<td>0.65</td>
<td>-0.13</td>
</tr>
<tr>
<td>W</td>
<td>-0.05</td>
<td>0.15</td>
<td>1</td>
<td>-0.05</td>
<td>0.01</td>
<td>-0.03</td>
<td>0.16</td>
<td>-0.14</td>
<td>0.1</td>
<td>-0.03</td>
</tr>
<tr>
<td>Y</td>
<td>-0.36</td>
<td>0.5</td>
<td>-0.05</td>
<td>1</td>
<td>0.18</td>
<td>-0.1</td>
<td>0.22</td>
<td>-0.04</td>
<td>0.41</td>
<td>-0.05</td>
</tr>
<tr>
<td>S</td>
<td>-0.19</td>
<td>0.19</td>
<td>0.01</td>
<td>0.18</td>
<td>1</td>
<td>-0.69</td>
<td>0.11</td>
<td>-0.05</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>F</td>
<td>0.16</td>
<td>-0.14</td>
<td>-0.03</td>
<td>-0.1</td>
<td>-0.69</td>
<td>1</td>
<td>-0.06</td>
<td>0.05</td>
<td>-0.2</td>
<td>0.04</td>
</tr>
<tr>
<td>nE</td>
<td>-0.19</td>
<td>0.33</td>
<td>0.16</td>
<td>0.22</td>
<td>0.11</td>
<td>-0.06</td>
<td>1</td>
<td>-0.09</td>
<td>0.24</td>
<td>-0.11</td>
</tr>
<tr>
<td>OF</td>
<td>0.02</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>-0.09</td>
<td>1</td>
<td>-0.04</td>
<td>0.12</td>
</tr>
<tr>
<td>nB</td>
<td>-0.72</td>
<td>0.65</td>
<td>0.1</td>
<td>0.41</td>
<td>0.24</td>
<td>-0.2</td>
<td>0.24</td>
<td>-0.04</td>
<td>1</td>
<td>-0.05</td>
</tr>
<tr>
<td>O</td>
<td>-0.05</td>
<td>-0.13</td>
<td>-0.03</td>
<td>-0.05</td>
<td>0.02</td>
<td>0.04</td>
<td>-0.11</td>
<td>0.12</td>
<td>-0.05</td>
<td>1</td>
</tr>
</tbody>
</table>

**Overlap**

**private attribute**

- M - Male
- F - Attr. Female
- W - White
- Y - Youth
- S - Smiling
- F - Frowning
- nE - No Eyewear
- OF - Obstr. Forehd.
- nB - No Beard
- O - Outdoors

**public attribute**
Conclusions

• We proposed a theoretical framework for quantitative balance between utility and privacy though policy based control of sensor data exposure.

• In our analysis we found promising results when we evaluated the PPLS algorithm in the context of automated face understanding.

• The algorithm we introduce is general, as it exploits the statistical properties of the data; and in the future it would be exciting to evaluate SensorSift in other sensor contexts.

• Available as Open Source!

   miro@cs.washington.edu
Thanks!

Liefeng  Xiaofeng  Jaeyeon  Yoshi

SecLab @ UW
Questions?

http://homes.cs.washington.edu/~miro/sensorSift