A Data-driven Synchronization Technique for Cyber-Physical Systems

Terrell R. Bennett, Nicholas Gans, Roozbeh Jafari

University of Texas at Dallas

April 13, 2015
Introduction

• The number of environmental sensors is rapidly growing
• Synchronization between the sensors in the system is critical for data fusion, estimation, and other signal processing
• Data between sensors may be unsynchronized due to
  – Inaccurate clock sources
  – Delays in time stamping
  – Wireless transmission delays
Motivation

• Homogeneity cannot be guaranteed in a Swarm of sensors
  – Wireless Data Transmission
  – Grid based power supply
  – Infrastructure/Environmental
• All sensors provide a data stream and related timestamps
• How can we use the events measured by the sensors to synchronize the sensor data
  – Locally Stored Data
  – Battery Powered
  – Mobile/Wearable
Contributions

• Offline technique for synchronizing sensor timing based on the sensor data
  – Use the physical world to synchronize data in the cyber world
  – Works with heterogeneous sensor data

• Methods to determine the physical or cyber events (i.e. couplings) in the sensor data streams that can be used for synchronization

• Methods to select a subset of the couplings to best improve the system synchronization
• Variety of clock generation circuits
• Typically trade off power for accuracy
  – Crystal Oscillators: ±20\textit{ppm}
  – Digitally Controlled Oscillators: ±5000\textit{ppm}
  – Voltage Controlled/Relaxation Oscillators: ±10000\textit{ppm}
• Oscillator inaccuracy and drift lead to clock errors and require synchronization
Sensor Clocking Test 1

• Single sensor with local storage
• ~22.5 hour run time
• Greater than 10 minutes of error over the collection
• Inconsistent clock due to battery
Sensor Clocking Test 2

• Sensor Test
  – 17 minutes of data
    • MSP430 Crystal
    • MSP430 DCO
    • MPU9150
  Relaxation
  Oscillator

~1 second delay
~25 second delay

17/04/15
Assumptions

• Difficult to control the quality of the clocks/time synchronization schemes in the presence several billions of sensors:
  – Some sensors may not have accurate clocks (due to power or cost constraints)
  – Some sensors may not be capable of supporting communication to enable time synchronization
  – Data from sensors cannot be effectively used if an acceptable level of time synchronization is not present
  – Data fusion from various sensors presents significant value that cannot be discarded
• Sensors $S_{1}$ and $S_{2}$ are making measurements of a physical phenomenon

• *Each sensor makes observations:* $o_{n} = \{x_{i_{n}}, t_{i_{n}}\}$
  – $n \in \{1, 2\}$ is the sensor number
  – $i \in \mathbb{N}$ is the observation number
  – $x_{i_{n}}$ and $t_{i_{n}}$ are the data and timestamp for the observation respectively
• **Definition:** An alignment point (AP) is a representation of a physical or cyber event in a sensor data stream that can be accurately distinguished and directly related to the same event in the data stream of another sensor (i.e. physical or cyber coupling)

\[
o_{i\uparrow k} \equiv o_{j\uparrow l} \quad \text{where } k \neq l
\]

\[
A = \{(o_{i\uparrow k}, o_{j\uparrow l}) \mid \exists i, k, j, l \text{ such that } o_{i\uparrow k} = o_{j\uparrow l}\}
\]

• We use the alignment points to correct the less accurate clock
Template Based AP Selection

• Synchronized templates of sensor interactions
  – Example:
Template Based AP Selection 2

• Use Dynamic Time Warping (DTW) to match the $S_{\downarrow 1}$ data stream
  – Matching not sensitive to speed variation
• Use Mutual Information to match the $S_{\downarrow 2}$ data stream to the $S_{\downarrow 1}$ data stream and the template

\[ I(X^{\uparrow k};X^{\uparrow l}) = \sum_{i} p(x^{\downarrow i^{\uparrow k}}, x^{\downarrow j^{\uparrow l}}) \log_2 \frac{p(x^{\downarrow i^{\uparrow k}}, x^{\downarrow j^{\uparrow l}})}{p(x^{\downarrow i^{\uparrow k}})p(x^{\downarrow j^{\uparrow l}})} \]
• The two sensors are measuring the same physical phenomena, but there is limited knowledge of the sensor data streams.

• Entropy is a measure of information content:
  \[ H(X^n) = -\sum_{i} p(x_i^n) \log_2 p(x_i^n) \]
  – The parts of the data with the highest entropy are considered useful for matching.
• Use peaks of the entropy calculation to determine segments of the $S\downarrow 1$ data to match against $S\downarrow 2$ data
• Use mutual information to match the data streams
• The peak of the mutual information calculations is the matching segment from $S\downarrow 2$
AP Subset Selection

- Entropy based AP selection can generate many possible segments for matching
- Select a subset of the APs $S \subseteq A$ for synchronization
- Two methods to select a subset of APs
  - Regional Peak Selection
  - Binary Search Selection
Regional Peak Selection

• Based on a quality score, $q_{\downarrow i}$, position, $p_{\downarrow i}$, and a region, $R$, select $S$

• Use the entropy value as the quality score, sort all alignment points, $a_{\downarrow i} \in A$
  
  $a_{\downarrow i} \in S$ iff $\max(q_{\downarrow i}), |p_{\downarrow i} - p_{\downarrow j}| > R$ and $i > j$

• Prioritizes quality while spreading alignments across the signal
Binary Search Selection

• Based on bisections of the data stream, find the entropy peak closest to the binary positions
• Select a level, \( L \), to generate up to \( 2^L - 1 \) points at positions, \( l \uparrow k \)
  \[ a \downarrow i \in S \text{ iff } |p \downarrow i - l \uparrow k| < |p \downarrow j - l \uparrow k|, \forall l \uparrow k \]
• Prioritizes equal spacing of alignment points across the data stream
Experiments

• Sensor
  – TI MSP430
  – Invensense MPU9150
  – Bluetooth data transmission

• Experiments
  – Template Based
  – Entropy Based
Template Based Experiments for two Sensor Synchronization

- Three experiments using templates

<table>
<thead>
<tr>
<th>Experiment Number</th>
<th>S↓1 location</th>
<th>S↓2 location</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Subject’s Right Thigh</td>
<td>Office Chair Arm</td>
</tr>
<tr>
<td>2</td>
<td>Subject’s Right Wrist</td>
<td>Office Door</td>
</tr>
<tr>
<td>3</td>
<td>Turntable Arm</td>
<td>Turntable Platter</td>
</tr>
</tbody>
</table>
Entropy Based Experiments

- Sensors measuring the same phenomenon
  - Stacked Sensors
  - Two Sensors on opposite arms of a chair
Metrics & Template Experiment Results

- **Metrics**
  - \( ppm = \Delta t / T \times 1,000,000 \)
  - \( E^{\downarrow n Tot} = \sum_i \nabla |t_\downarrow - t_\downarrow| / n \)

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Thigh, Chair</th>
<th>Wrist, Door</th>
<th>Turntable</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E^{\downarrow n Tot} ) (ms)</td>
<td>72.8</td>
<td>51.6</td>
<td>22.9</td>
<td>13.3</td>
</tr>
<tr>
<td>( ppm )</td>
<td>5,131.1</td>
<td>1,862.2</td>
<td>199.9</td>
<td>11,907</td>
</tr>
</tbody>
</table>

![Graph showing acceleration over time for different experiments and sensor data types.]
## Entropy Experiment Results

### Non-Template Binary Search Method

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Stacked</th>
<th>Two on Chair</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_{\text{tot}} ) (ms)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{min} )</td>
<td>1.2</td>
<td>6.0</td>
<td>10.9</td>
</tr>
<tr>
<td>( \text{max} )</td>
<td>16.4</td>
<td>24.9</td>
<td>216.1</td>
</tr>
<tr>
<td>( \text{avg} )</td>
<td>5.9</td>
<td>11.4</td>
<td>70.2</td>
</tr>
<tr>
<td>ppm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{min} )</td>
<td>35.3</td>
<td>291.8</td>
<td>11,443</td>
</tr>
<tr>
<td>( \text{max} )</td>
<td>423.5</td>
<td>513.0</td>
<td>14,259</td>
</tr>
<tr>
<td>( \text{avg} )</td>
<td>195.8</td>
<td>359.2</td>
<td>12,224</td>
</tr>
</tbody>
</table>

### Non-Template Region Peak Method

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Stacked</th>
<th>Two on Chair</th>
<th>Original</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E_{\text{tot}} ) (ms)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{min} )</td>
<td>2.9</td>
<td>10.4</td>
<td>10.9</td>
</tr>
<tr>
<td>( \text{max} )</td>
<td>27.4</td>
<td>33.2</td>
<td>216.1</td>
</tr>
<tr>
<td>( \text{avg} )</td>
<td>7.4</td>
<td>27.3</td>
<td>70.2</td>
</tr>
<tr>
<td>ppm</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \text{min} )</td>
<td>107</td>
<td>215.7</td>
<td>11,443</td>
</tr>
<tr>
<td>( \text{max} )</td>
<td>529.1</td>
<td>838.9</td>
<td>14,259</td>
</tr>
<tr>
<td>( \text{avg} )</td>
<td>208.5</td>
<td>680.7</td>
<td>12,224</td>
</tr>
</tbody>
</table>

### Diagram

- **Gold Standard vs Input Clock**
- **Synchronized Data vs Input Clock**

### Graph

- Error (ms) vs Samples

\[ 0 \leq \text{Error (ms)} \leq 30 \]

\[ 0 \leq \text{Samples} \leq 14000 \]
Conclusions

• Sensor-rich environments are likely to have heterogeneous sensors
• Synchronization increases the value of data received from these systems
• An offline synchronization technique can be beneficial for synchronizing TerraSwarm data
• Cyber and physical couplings can be used to synchronize the data in an offline manner
Thanks & Questions