TerraSwarm

A Data-driven Synchronization Technique for Cyber-Physical Systems

Terrell R. Bennett, Nicholas Gans, Roozbeh Jafari

University of Texas at Dallas

April 13, 2015

Sponsored by the TerraSwarm Research Center, one of six centers administered by the STARnet phase of the Focus Center Research Program (FCRP) a Semiconductor Research Corporation program sponsored by MARCO and DARPA.

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

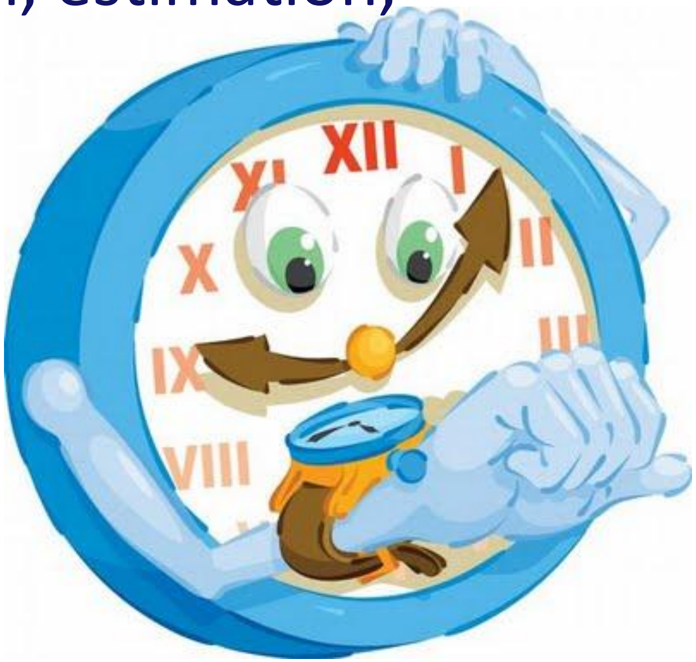


Image from http://blog.onlineclock.net/

Motivation

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

Image from www.ti.com

Contributions

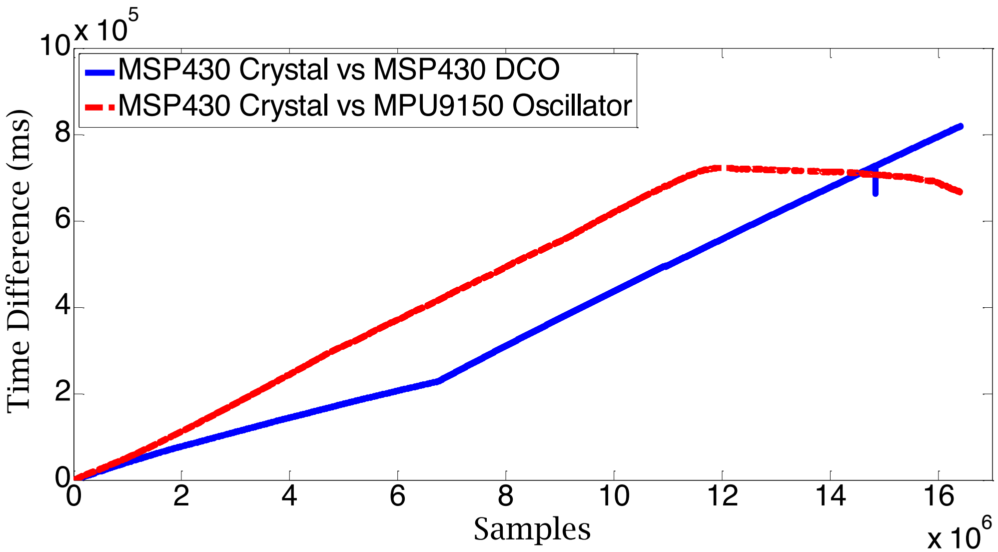
- Offline technique for synchronizing sensor timing based on the sensor data
  - Use the physical world to synchronize data in the cyber world
  - Works with heterogeneous sensor data
- Methods to determine the physical or cyber events (i.e. couplings) in the sensor data streams that can be used for synchronization
- Methods to select a subset of the couplings to best improve the system synchronization

Background

- Variety of clock generation circuits
- Typically trade off power for accuracy
  - Crystal Oscillators: $\pm 20ppm$
  - Digitally Controlled Oscillators: $\pm 5000ppm$
  - Voltage Controlled/Relaxation Oscillators: $\pm 10000ppm$
- Oscillator inaccuracy and drift lead to clock errors and require synchronization

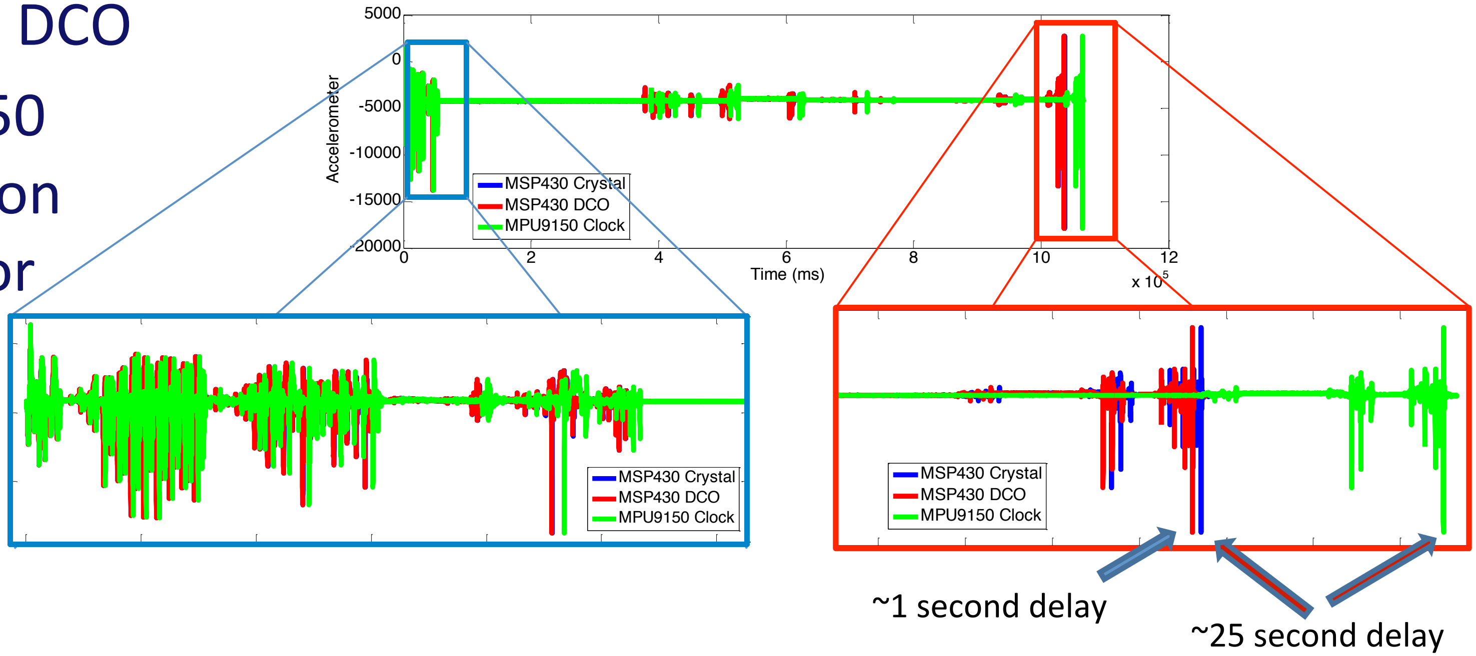
Sensor Clocking Test 1

- 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



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

Two Sensor Alignment Formulation

- Sensors S_1 and S_2 are making measurements of a physical phenomenon
- *Each sensor makes observations*: $o_i^n = \{x_i^n, t_i^n\}$
  - $n \in \{1,2\}$ 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

Alignment Points

- **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)

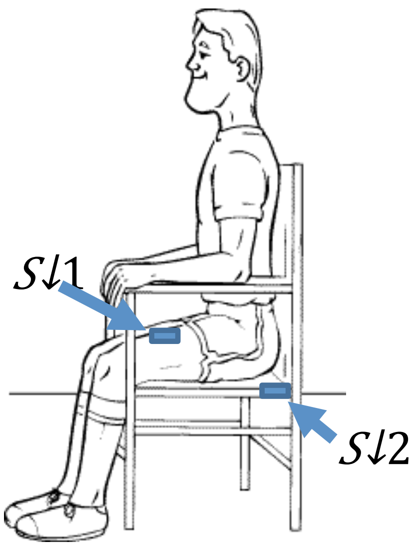
$$o_i^k \equiv o_j^l \text{ where } k \neq l$$

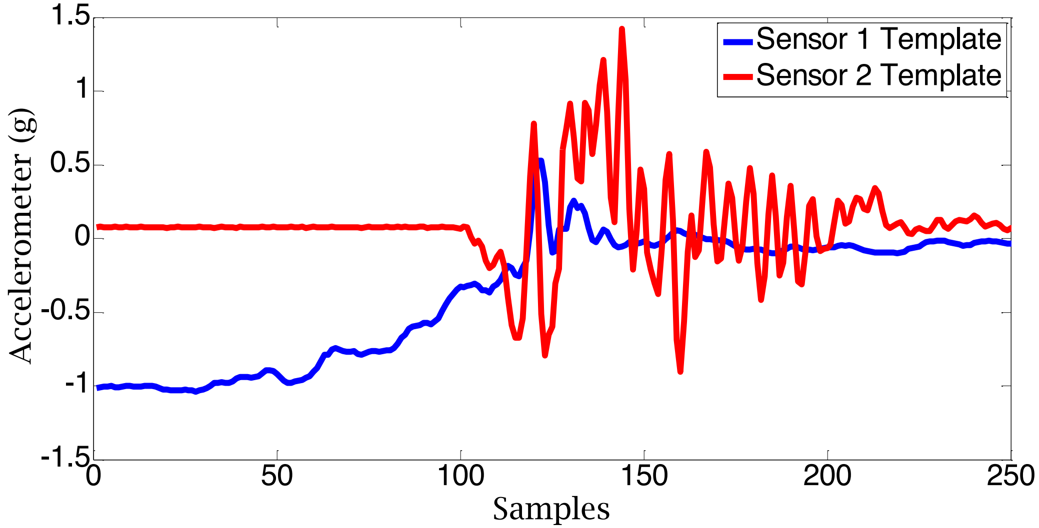
$$\boldsymbol{A} = \{(o_i^k, o_j^l) \mid \exists\, i,k,j,l \text{ such that } o_i^k \equiv o_j^l\}$$

- We use the alignment points to correct the less accurate clock

Template Based AP Selection

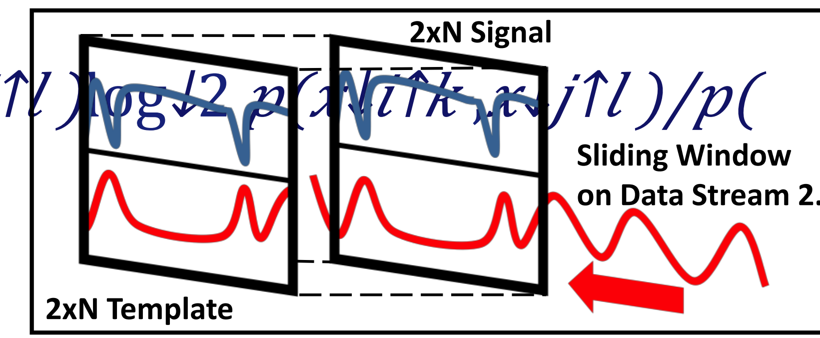
- Synchronized templates of sensor interactions
  - Example:





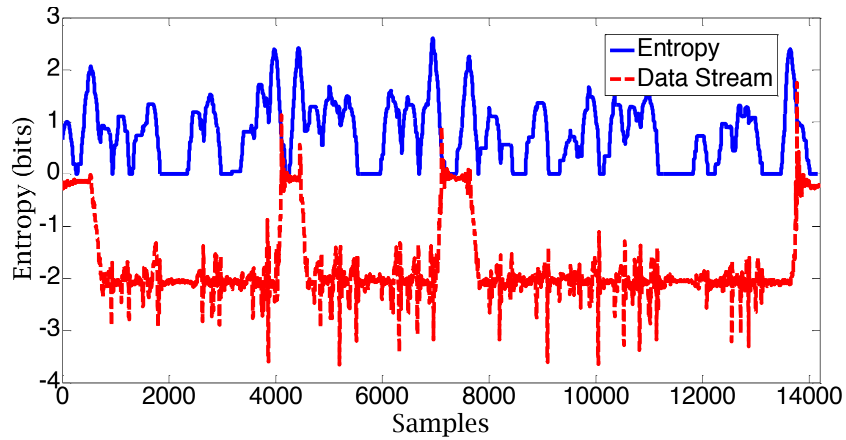
Template Based AP Selection 2

- Use Dynamic Time Warping (DTW) to match the S_1 data stream
  - Matching not sensitive to speed variation
- Use Mutual Information to match the S_2 data stream to the S_1 data stream and the template
- $I(X^k;X^l)=\sum p(x_i^k,x_j^l)\log_2 p(x_i^k,x_j^l)/p(x_i^k)p(x_j^l)$



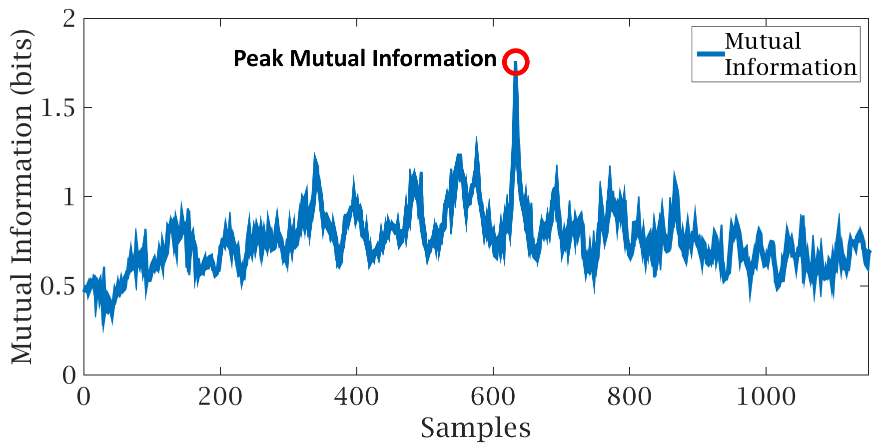
Entropy Based AP Selection

- 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
  - $H(X^n) = -\sum p(x_i^n) \log_2 p(x_i^n)$
  - The parts of the data with the highest entropy are considered useful for matching



Entropy Based AP Selection 2

- Use peaks of the entropy calculation to determine segments of the S_1 data to match against S_2 data
- Use mutual information to match the data streams
- The peak of the mutual information calculations is the matching segment from S_2



AP Subset Selection

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

Regional Peak Selection

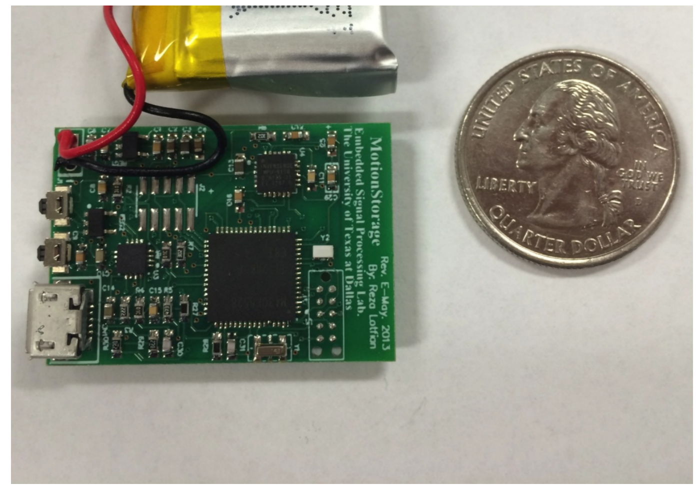
- Based on a quality score, q_i, position, p_i, and a region, R, select $\boldsymbol{S}$
- Use the entropy value as the quality score, sort all alignment points, $a_i \in \boldsymbol{A}$
  - $a_i \in \mathbf{S}$ iff $\max(q_i)$, $|p_i - p_j| > R$ and $i > j$
- Prioritizes quality while spreading alignments 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_k
  - $a_i \in \mathbf{S}$ iff $|p_i - l_k| < |p_j - l_k|, \forall\ l_k$
- Prioritizes equal spacing of alignment points across the data stream

Experiments

- Sensor
  - TI MSP430
  - Invensense MPU9150
  - Bluetooth data transmission

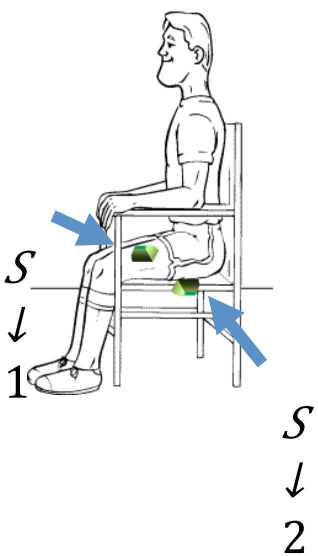


- Experiments
  - Template Based
  - Entropy Based

Template Based Experiments for two Sensor Synchronization

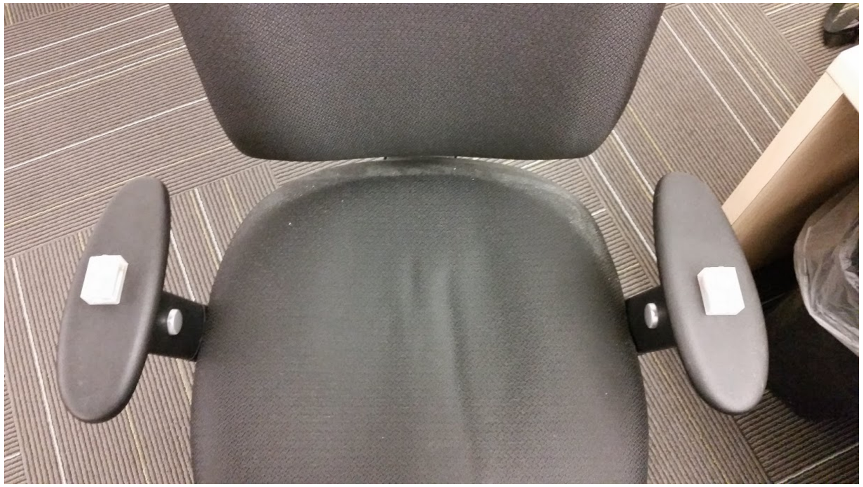
- Three experiments using templates

Experiment Number	S_1 location	S_2 location
1	Subject’s Right Thigh	Office Chair Arm
2	Subject’s Right Wrist	Office Door
3	Turntable Arm	Turntable Platter



Entropy Based Experiments

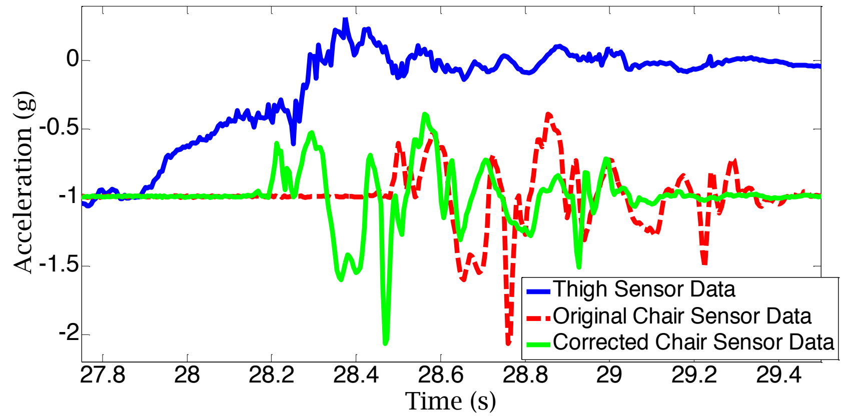
- Sensors measuring the same phenomenon
  - Stacked Sensors
  - Two Sensors on opposite arms of a chair



Metrics & Template Experiment Results

- Metrics
  - $ppm = \Delta t / T \times 1{,}000{,}000$
  - $E_{nTot} = \sum_{i}^{n} |t_i - t_i| / n$

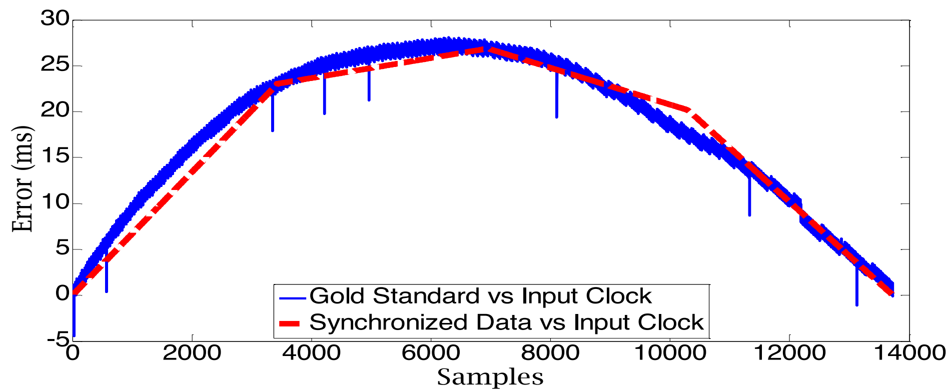
Template Based Method				
Experiment	Thigh, Chair	Wrist, Door	Turn-table	Original
E_{nTot} (ms)	72.8	51.6	22.9	13.3
ppm	5,131.1	1,862.2	199.9	11,907



Entropy Experiment Results

Non-Template Binary Search Method				
Experiment		Stacked	Two on Chair	Original
***E_{nTot}* (ms)**	***min***	1.2	6.0	10.9
	max	16.4	24.9	216.1
	avg	5.9	11.4	70.2
ppm	***min***	35.3	291.8	11,443
	max	423.5	513.0	14,259
	avg	195.8	359.2	12,224

Non-Template Region Peak Method				
Experiment		Stacked	Two on Chair	Original
***E_{nTot}* (ms)**	***min***	2.9	10.4	10.9
	max	27.4	33.2	216.1
	avg	7.4	27.3	70.2
ppm	***min***	107	215.7	11,443
	max	529.1	838.9	14,259
	avg	208.5	680.7	12,224



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