EXPERIMENTS AND ANALYSIS OF QUALITY- AND ENERGY-AWARE DATA AGGREGATION APPROACHES IN WSNS

Cinzia Cappiello – Fabio A. Schreiber

Politecnico di Milano
Dipartimento di Elettronica ed Informazione
TECHNOLOGICAL ISSUES WITH SENSOR NODES IN WSNs

- **MEMORY**
  - SMALL ALGORITHM FOOTPRINT
  - SMALL DATA BUFFERS
    - STORE FEW DATA
    - FOR A LIMITED TIME SPAN
  - TRANSFER DATA ELSEWHERE AS SOON AS POSSIBLE

- **POWER**
  - BATTERY LIFE IS LIMITED
  - TRANSMISSION IS THE MOST POWER CONSUMING FUNCTION
  - LIMIT TRANSMISSION FREQUENCY AND DURATION

- **PRESERVE DATA QUALITY**
  - ACCURACY
  - PRECISION
  - TIMELINESS
DATA QUALITY DIMENSIONS

- **ACCURACY**
  
  the degree of conformity of a measured or computed quantity to its actual (true) value
  
  \[ |v_{avg} - v_{ref}| < \varepsilon_{acc} \]

- **PRECISION**
  
  the degree to which repeated measurement show the same or similar results
  
  (small variance)
  
  \[ 1/n^*\sum_{n=1}^{N} (v_{n} - \mu)^2 < \varepsilon_{prec} \]

- **TIMELINESS**
  
  - **CURRENCY**
    
    the time interval from the instant the value was sampled to the instant at which it is sent to the base station

  - **VOLATILITY**
    
    the amount of time during which data remain valid

  \[ \text{Timeliness} = \max(1 - \text{Currency/Volatility}; 0) \]

F. A. Schreiber, C. Cappiello

Energy-QoD
DATA COMPRESSION IS THE PROCESS OF ENCODING INFORMATION USING FEWER BITS

- Traditional data compression algorithms are not feasible for WSNs owning to their size and complexity

- Data compression algorithms proposed for WSN are focused on:
  - Temporal correlation/aggregation
  - Spatial correlation/aggregation

- … and they are characterized by the following limitations:
  - Specialization on specific contexts
  - Utilization on predictive models and thus, limitations on acceptable values
  - Data Quality is scarcely considered
  - Outliers are discarded
SYSTEM STRUCTURE

INPUT BUFFER 1

INPUT BUFFER 2

COMPRESSION

OUTPUT BUFFER
BASIC PRINCIPLES OF THE PROPOSED AGGREGATION ALGORITHM

- **Accuracy** is represented by the window height.
- Values falling **within** the window can be considered similar enough to be fairly represented by their average.
- Values falling **outside** the window are **outliers**.
- Outliers can be **occasional** or **consecutive**: in any case, outliers information must be preserved for further investigation.

F. A. Schreiber, C. Cappiello  Energy-QoD
CONSIDERED CASES

- **EXPECTED TREND**
  - The trend is regular
  - Values are both precise and accurate

- **SLOW CHANGE**
  - The trend shows an unpredicted but lasting variation
  - Values are still precise, but not accurate

- **OSCILLATORY / BURSTY / SUDDEN CHANGE**
  - The trend shows discontinuous variations
  - Values are not precise, but they can be both accurate or not

Any data stream can be described as a combination of these cases.
By considering $Z$ aggregate values and $J$ outliers out of a set of $N$ measures, the algorithm is considered efficient if the output is composed by $(Z+J)$ values instead of $N$ where $(Z+J)<<N$.
OUTLIER DETECTION

- VALUES NOT ACCURATE BUT PRECISE
  - INCOMING VALUES DON’T MATCH THE REFERENCE VALUE
  - THEIR STANDARD DEVIATION IS SMALL

CHANGE IN THE BEHAVIOUR

- VALUES STILL ACCURATE AND PRECISE

ERROR

F. A. Schreiber, C. Cappiello
Energy-QoD
ALGORITHM BANDWIDTH

- COMPRESSING DATA AMOUNTS TO LOWERING THE BANDWIDTH OF THE MEASUREMENT SYSTEM

- THE WINDOW WIDTH DETERMINES THE NUMBER OF MEASURED VALUES WHICH ARE AGGREGATED
  - 1 POINT WINDOW → NO COMPRESSION → MAX BANDWIDTH

- THE WINDOW WIDTH ALSO DETERMINES THE TIMELINESS BY WHICH DATA ARE DELIVERED TO THE BASE STATION
ALGORITHM BANDWIDTH

- The window width must be adapted to catch very steep/sudden changes

![Diagram showing window width adjustment for steep changes](image)
ALGORITHM INPUT/OUTPUT

INPUT PARAMETERS
- TIME SERIES
- EXPECTED VALUE
- ACCURACY TOLERANCE
- PRECISION TOLERANCE
- WINDOW WIDTH
- CONTINUITY INTERVAL

V = <v₁, v₂, … vₙ>

v_{ref}

\varepsilon_{acc}

\varepsilon_{prec}

N

C

OUTPUT PARAMETERS
- AGGREGATE VALUES
- OUTLIERS

T = <a₁,t₁>; <a₂,t₂>; … <aₙ,tₙ>

O = <o₁,t₁>; <o₂,t₂>; … <oᵢ,tᵢ>

ALGORITHM COMPLEXITY
O(N)

ALGORITHM FOOTPRINT
11 KB RAM; 1 KB ROM
EXPERIMENTAL SET UP

\[ 100 \ \Omega \leq Z_R(t) \leq 1000 \ \Omega \] (measured)

\[ R = 1 \ \Omega \rightarrow R + Z_R \approx Z_R \]

\[ 0 \ mA \leq i \leq 30 \ mA \] (Data sheet)

\[ 0 \ mV \leq \Delta V \leq 30 \ mV \]

\[ E(t) = \int_0^t \left[ \frac{v_2(t')}{R} \right] * v_1(t') \ dt' \]
ALGORITHM BEHAVIOUR

WITH AGGREGATION

WITHOUT AGGREGATION (BYPASS)

7 TRANSMITTED VALUES, 30mJ

60 TRANSMITTED VALUES, 120mJ

70% ENERGY SAVINGS

F. A. Schreiber, C. Cappiello

Energy-QoD
ALGORITHMS COMPARISON: CONSIDERED SIGNALS

(a) $C_2N_2$ absorption spectrum

(b) $C_2N_2$ absorption spectrum FM

F. A. Schreiber, C. Cappiello  
Energy-QoD
The data sets shown in the previous slide have been processed by using the algorithm proposed and two other aggregation algorithms:


The comparison among algorithms have been based on three main criteria:

- **Compression rate**: the degree with which data have been aggregated.

- **Energy savings**: the degree with which the aggregation allows sensors to save energy with respect to the case in which all the original values are sent to the base station.

- **Correctness**: the degree with which the aggregated data allow the base station to retrieve the original trend. Correctness has been evaluated by using the Mean Absolute Error (MAE) and the related Mean Absolute Percentage Error (MA%E).
DATA SET (A) RESULTS
DATA SET (A) RESULTS
DATA SET (A) RESULTS

Compression rate

Energy Reduction

MAE in case of non linear trends

<table>
<thead>
<tr>
<th></th>
<th>Authors</th>
<th>Schoellhammer et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>0.0008</td>
<td>0.0009</td>
</tr>
<tr>
<td>(b)</td>
<td>0.0011</td>
<td>0.0014</td>
</tr>
<tr>
<td>(c)</td>
<td>0.0008</td>
<td>0.0009</td>
</tr>
</tbody>
</table>
DATA SET (B) RESULTS

Systematic error due to the processing time shift
DATA SET (B) RESULTS

F. A. Schreiber, C. Cappiello
Energy-QoD
DATA SET (B) RESULTS

F. A. Schreiber, C. Cappiello

Energy-QoD

Input data set

Schoellhammer et al.
DATA SET (B) RESULTS

Compression rate

Energy Savings

F. A. Schreiber, C. Cappiello  Energy-QoD
SUMMARY COMPARISON AND CONCLUSIONS

- No single algorithm is «the best»
- Transmission procedures with packed based protocols can affect the analysis
  - Higher packing factors improve energy efficiency
  - Higher transmission delays negatively affect timeliness
- Adaptable procedures should be used on the basis of
  - The peculiar features of the signals to be processed
  - The quality requirements of the applications

<table>
<thead>
<tr>
<th>Authors</th>
<th>Compression rate/Energy saving</th>
<th>MA%E/Compression rate</th>
<th>MA%E/Energy saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lazaridis et al.</td>
<td>2.5</td>
<td>0.64</td>
<td>1.6</td>
</tr>
<tr>
<td>Schoellhammer et al.</td>
<td>1.55</td>
<td>1.77</td>
<td>2.77</td>
</tr>
<tr>
<td></td>
<td>2.84</td>
<td>0.615</td>
<td>1.75</td>
</tr>
</tbody>
</table>